Approach for Improved Signal-Based Fault Diagnosis of Hot Rolling Mills

Von der Fakultät für Ingenieurwissenschaften,
Abteilung Maschinenbau und Verfahrenstechnik,
der
Universität Duisburg-Essen
zur Erlangung des akademischen Grades
einer
Doktorin der Ingenieurwissenschaften
Dr.-Ing.
genehmigte Dissertation

von

Astrid Rother

aus

Krefeld

Gutachter: Univ.-Prof. Dr.-Ing. Dirk Söffker
Prof. Dr.-Ing. Mohieddine Jelali, Priv. Doz.
Univ.-Prof. Dr. rer. nat. Johannes Gottschling
Tag der mündlichen Prüfung: 20. Januar 2016

Acknowledgment

Sincere thanks to Univ.-Prof. Dr.-Ing. D. Söffker and Prof. Dr.-Ing. M. Jelali for the continuous support of this project and the willingness to examine my thesis. Likewise I thank Univ.-Prof. Dr. rer. nat. J. Gottschling for his effort in examining my thesis.

I would like to express my gratitude to the management of ThyssenKrupp Steel Europe AG, the former director of the hot strip mill Bochum Dipl.-Ing. E.-A. Becker, and the director of the hot strip mill Bochum Dr.-Ing. C. Evers who gave me the opportunity to do research in an industiral environment.

I am grateful to my colleagues at ThyssenKrupp Steel Europe AG, hot strip mill Bochum for their helpful support. Especially I would like to thank Dr.-Ing. I. Jäckel, Dipl.-Ing. T. Pulcher, Dipl.-Ing. P.Hoy, and Dipl.-Ing. B. Röttgers for the provision of hardware and numerous inspiring discussions.

Duisburg, January 2016 Astrid Rother

Abstract

The approach introduced here is able to detect two specific severe faults, to identify them, to distinguish between four different system states, and to give a prognosis on the system behavior. The presented work investigates the condition monitoring of the complex production process of a hot strip rolling mill. A signal-based fault diagnosis and fault prognosis approach for strip travel is developed. A literature review gives an overview about previous research on related topics. It is shown that the great amount of previous work does not cope with the problems treated in this work and that further investigation is necessary to provide a satisfactory solution. The design of a new signal processing chain is presented and the signal processing steps are detailed. The classification task is differentiated into fault detection, fault identification and fault prognosis. The proposed approach combines five different methods for feature extraction, namely short time Fourier transform, continuous wavelet transform, discrete wavelet transform, Wigner-Ville distribution, and empirical mode decomposition, with two different classification algorithms, namely support vector machine and a variation of cross-correlation, the latter developed in this work. Combinations of these feature extraction and classification methods are applied to rolling force data originating from a hot strip mill.

Kurzfassung

Der hier vorgestellte Ansatz ist in der Lage, zwei spezifische schwere Fehler zu erkennen, sie zu identifizieren, zwischen vier verschiedenen Systemzuständen zu unterscheiden und eine Prognose bezüglich des Systemverhaltens zu geben. Die vorliegende Arbeit untersucht die Zustandsüberwachung des komplexen Herstellungsprozesses eines Warmbandwalzwerks. Eine signalbasierte Fehlerdiagnose und ein Fehlerprognoseansatz für den Bandlauf werden entwickelt. Eine Literaturübersicht gibt einen Überblick über die bisherige Forschung zu verwandten Themen. Es wird gezeigt, dass die große Anzahl vorheriger Arbeiten diese Thematik nicht gelöst hat und dass weitere Untersuchungen erforderlich sind, um eine zufriedenstellende Lösung der behandelten Probleme zu erhalten. Die Entwicklung einer neuen Signalverarbeitungskette und die Signalverarbeitungsschritte sind detailliert dargestellt. Die Klassifikationsaufgabe wird in Fehlerdiagnose, Fehleridentifikation und Fehlerprognose differenziert. Der vorgeschlagene Ansatz kombiniert fünf verschiedene Methoden zur Merkmalsextraktion, nämlich Short-Time Fourier Transformation, kontinuierliche Wavelet Transformation, diskrete Wavelet Transformation, Wigner-Ville Distribution und Empirical Mode Decomposition, mit zwei verschiedenen Klassifikationsalgorithmen, nämlich Support-Vektor Maschine und eine Variation der Kreuzkorrelation, wobei letztere in dieser Arbeit entwickelt wurde. Kombinationen dieser Merkmalsextraktion und Klassifikationsverfahren werden an Walzkraft-Daten aus einer Warmbreitbandstraße angewendet.

Contents

List of Figures

List of Tables

List of acronyms

AGC	Automatic Gauge Control
ARMA	Autoregressive Moving Average
CC	Cross Correlation
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
FFT	Fast Fourier Transform
FMEA	Failure Mode and Effect Analysis
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FT	Fourier Transform
FWT	Fast Wavelet Transform
GRNN	General Regression Neural Network
HHT	Hilbert-Huang Transform
HT	Hilbert Transform
IFFT	Inverse Fast Fourier Trasform
IMF	Intrinsic Mode Function
ISOMAP	Isometric Feature Mapping
LMD	Local Mean Decomposition

MB	Model-based
MCA	Morphological Component Analysis
MW	Mother Wavelet
NN	Neural Networks
ROC	Receiver Operating Characteristic
SALSA	Split Augmented Lagrangian Shrinkage Algorithm
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
SWT	Stationary Wavelet Transform
TKSE	ThyssenKrupp Steel Europe AG
TN	True Negative
TP	True Positive
TPR	True Positive Rate
TQWT	Tunable Q-factor Wavelet Transform
WPT	Wavelet Packet Transform
WT	Wavelet Transform
WVD	Wigner Ville Distribution

1 Introduction

1.1 Motivation and task of this research

Rolling is an important processing method in the metal industry. Strips, plates, and sheets from hot and cold rolling mills are widely used in industrial processes. From the customer point of view, slightest changes in geometry, surface condition, and thickness may spoil a strip. Therefore, the necessary accuracy of product dimension (strip thickness and flatness) is in the micrometer range. Especially for the automotive industry, the requirements on surface quality (roughness) are high. Customer demands in product quality lead to increased process complexity. The variety of possible damages increases.

The general interest of the plant operator is therefore related to the reduction of down-times, an increase in product quality, a useful lifetime extension of machines and machine parts. Profound knowledge about the production plant and the exact system state is needed to cope these demands. Figure 1.1 illustrates the hot strip rolling process. The rolling stands of a finishing mill are in the background. An orange-whitish glowing metal strip is passing the mill. In the foreground, a set of seven spare roll pairs is prepared for the next exchange of work rolls.

The task of system diagnosis can be mastered based on reliable knowledge of the system's condition resulting from continuous monitoring of the machine state. To detect faults in process operation, condition monitoring can be used. Here, the system state fault is defined as a state or behavior out of given or defined parameters. Consequently, fault symptoms can be detected as those deviations from regular process behavior.

In the last decades, various condition monitoring systems and approaches emerged. They are commonly used in diverse industrial areas to detect, diagnose, and analyze the deterioration of system performance [1, 2, 3, 4, 5]. Analyzing a plant's condition starts with the extraction of information embedded in specific signals. To gain such signals, different measurement principles are available, e.g. optical, acoustical, mechanical, and combinations of those [6].

Figure 1.1: Finishing mill of a hot strip rolling mill [7]

The obtained signal may differ, especially if diverse sensors are used. Depending on the analysis method, it can be difficult to capture the necessary information. Due to occasionally rough conditions, it may be impossible to capture the aimed signal with the necessary quality for the next processing steps (like classification). Feng et al. [8] give an overview of the application of condition monitoring in condition-based maintenance. Jardine et al. [9] review the three layers of condition-based maintenance: data acquisition, data processing, and analysis. The authors describe the use of models, algorithms, and techniques aiming at the scope of maintenance decision support (diagnosis/ prognosis), emphasizing that both, event data and condition monitoring data, are important.

To reach these goals, several different approaches can be used. The three main strategies for condition monitoring and prognosis are according to Lee et al. [10]: model-based, data-driven, and hybrid. Ma et al. [11] further differentiate between data-driven and

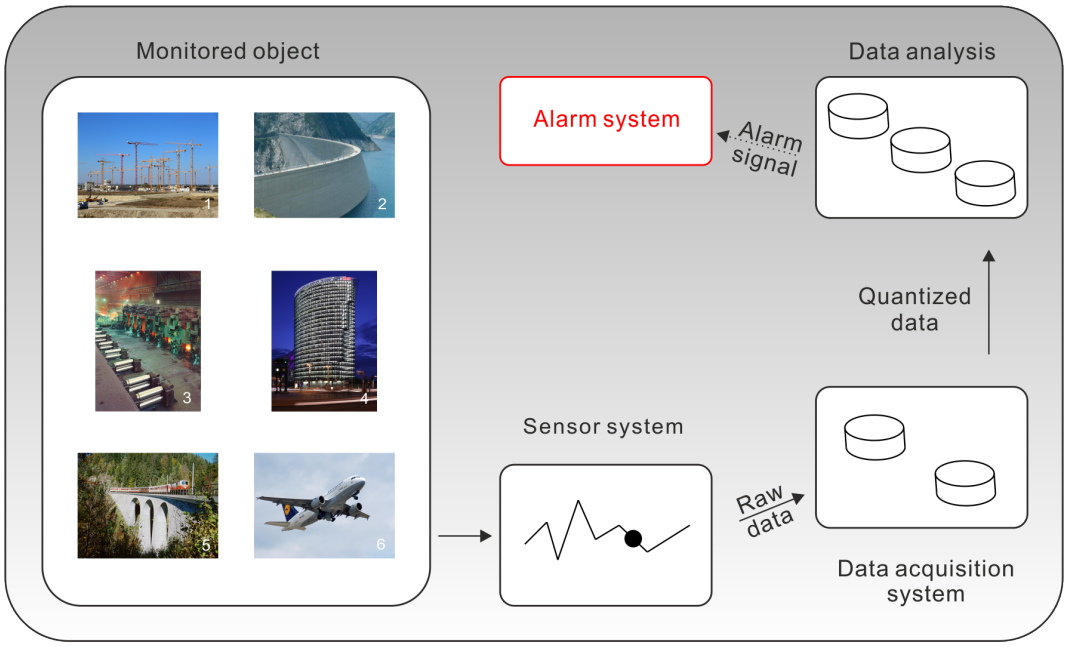


Figure 1.2: General signal-based diagnosis and prognosis concept (photos taken from [14], [15], [7], [16], [17], [18] in numerical order)

signal-based methods, in the way that data-driven approaches use more complex analysis methods to allow fault detection and isolation tasks. The development of a complete mathematical model of the considered system and the estimation of the parameters that depict the actual behavior is usually costly. Isermann [1] gives an overview on different modeling approaches. Samy et al. [3] compare approaches using a physical model to methods not using a physical model. An example of modeling and parameter estimation is given by Lal and Tiwari [4]. Hou and Wang [5] prefer data-driven approaches, that use empirical knowledge about the common case and are able to detect deviations. The authors claim that since no analytical a priori knowledge is needed, this class of approaches is less costly and less time consuming compared to model-based approaches. A fusion of data-driven and model-based methods leads to hybrid models. A hybrid approach combines the output of both strategies and therefore aims to benefit from the advantages [12, 13].

So far, signal-based and data-driven approaches in strip rolling mills have received very little attention in literature. This work presents developments in the use of signal-based approaches for system monitoring, in detail for fault detection, fault identification, and fault prognosis. Figure 1.2 shows a generalized signal-based diagnosis concept. Goal is to improve the use of operational information sources for in-time fault prediction.

To enhance the informative value of operative measured signals, different methods for feature extraction and classification are applied. The online applicability in an industrial area is an additional interest of the approaches. In many production lines, the responsibility for the integrity of the process lies in the hand of the operators. The information given to them by the applied analysis methods should be precise and comprehensible to allow immediate decisions. Therefore, even small improvements to an existing system can be a progress.

The contributions on time-frequency-based signal processing in dynamic processes are reviewed. Special emphasis is given to condition monitoring applied to strip rolling mills.

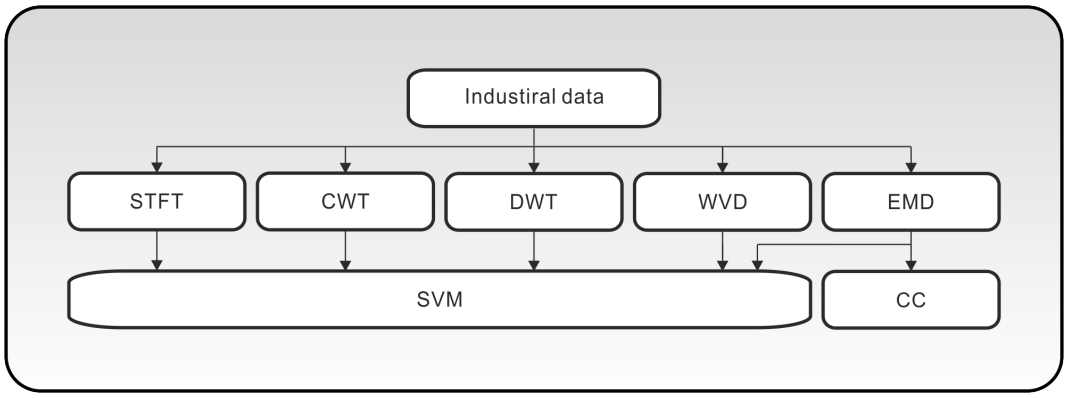


Figure 1.3: Overview on selected methods of condition monitoring

The selected methods for feature extraction and classification are shown in Figure 1.3, namely Short Time Fourier Tranform (STFT), Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), Wigner Ville Distribution (WVD), and Empirical Mode Decomposition (EMD) for feature extraction and Support Vector Machine (SVM) and Cross Correlation (CC) for classification. The suitability of the method combinations for fault detection, fault identification, and fault prognosis is evaluated.

1.2 Scientific contribution and structure of the thesis

In this work, a review of approaches related and applied to hot strip mills is given for the first time [19], and the corresponding results are discussed. Signal-based fault detection, fault identification, and fault prognosis strategies for an industrial application are developed. The performance of four renown methods in feature extraction applied to real process data from a hot strip mill combined with two classifiers is evaluated. The results show differences in the potential of the methods when applied to real process data.

The development of a new combination of feature extraction and classification is shown. Feature extraction of the new monitoring approach is accomplished by EMD, and classification by thresholding of CC with good performance in fault detection, identification, and prognosis. The relative computational load and response time are reduced by a new decision making method. In-time shutdown can avoid severe damages on the product and the production line. Information at an early stage is of great importance for alert triggering. The prognosis method presented in this work is able to indicate an upcoming fault early.

This thesis is divided into six chapters. In Chapter 1, the motivation for the work performed in this thesis is given. Chapter 2 presents a literature review on recent developments in the field of signal-based diagnosis in hot strip mills. In Chapter 3, the application site is introduced. A short background on rolling is given and the target deviations in strip travel are described. Chapter 4 details the approach taken and the mathematical tools used in the analysis and describes the methodology used in the design of the experiments. In Chapter 5, the experimental results are presented, and the significance of the results is discussed. Chapter 6 summarizes the main conclusions of this thesis and presents an outlook for future work.

2 Literature review

The complexity of control systems and operator activities has increased in the last decades. To reduce the necessity of expert-knowledge in every day use, the degree of automation is expanded, and automized diagnosis and prognosis become more important tasks in modern industrial plants [1, 2, 3, 4, 5]. The objective is to ensure the structural safety, to extend the life time of components, and to determine severe failures before they appear. One approach is the continuous surveillance of industrial machines to gather information on the system condition. This is common practice in production lines [9]. As Julcher [20] presents, signal-based analysis methods can well be used for fault diagnosis in industrial plants.

This chapter presents the state of the art and an overview of selected signal-based analysis methods. So far, the applications to hot strip mills concentrate on models and time-based or frequency- based analysis. To distinguish the presented approach from the usual practice, a time-frequency-based approach is used. It will be shown that time-frequency-based analysis methods rarely found application in the field of hot strip mills. Based on the evaluation of signal-based analysis methods applied to comparable industrial environments, five analysis methods are chosen for further investigation. The content of this chapter is based on contributions already published [19].

2.1 General applications of signal-based analysis

In general, signal-based analysis is done in the time domain, the frequency domain, or the time-frequency domain [8]. In industrial practice, time-based analysis of signals is well established in condition monitoring [21]. Intermittency, trend monitoring, threshold monitoring of statistical characteristics like mean, peak, standard derivation, and root mean square are common techniques for qualitative fault analysis. As a variation, Caesarendra et al. [22] base their condition monitoring approach on circular domain features and claim superiority to time-frequency analysis methods. Serido et al. [23] use autoregressive moving average-based methods (ARMA), a more elaborate time-based analysis.

Nandi et al. [24] extend the treatment of vibration signal to kurtosis.

Frequency-based analysis allows to handle dynamic attributes, namely for application in rotating machinery with periodic signals. The most conventional frequency-based analysis is fast Fourier transform (FFT). The FFT shows frequency components with amplitudes sufficient to protrude the noise. To extract impacts with very low energy and therefore hidden features, frequency filters, envelope analysis, side band structure analysis, Hilbert transform, power spectrum, cepstrum, matched filter (signal to noise ratio) and others are used.

The mathematical prerequisites for frequency domain analysis do not allow the handling of non-stationary, non-periodic occurrences. To gain event-related and time-related information from such signals, it is recommendable to analyze via the time-frequency domain. Using this class of methods, periodicity is no longer demanded and singular time events can be handled. The methods will give information on the occurrence time of events.

A recent overview on the application fields and the benefits and drawbacks of different fault diagnosis methods is given by Lee et al. [10]. It is stated that no systematic method for the development of a health monitoring system exists. The authors give a summary of health monitoring tools applied to five critical components. Table 2.1 gives a new interpretation of their contribution. The table shows the application fields of the methods presented by Lee et al. From this table, it can be seen that some methods are mainly applied to one application field while others are in widespread use, e.g. the wavelet transform. Some of the methods show only a small number of applications, e.g. the Wigner-Ville distribution. For the approaches given in table 2.1, commonly analyzed measurements in matters of bearings and gears are vibration, oil debris, and acoustic emission. In the field of shaft monitoring, vibration is used. Pump analysis focuses on vibration, pressure, and acoustic emission. The monitoring of generators is based on stator current, stator voltage, magnetic fields, and frame vibrations. The characteristics are to be summarized as follows.

Time domain analysis [25] regards the waveform of an input signal, e.g. the comparison of two wave forms. It does not provide further information. Fourier transform [26] represents the frequency components of an input signal, but it is limited to periodic signals and has a lack of information on the occurrence time. Short time Fourier transform [27, 28] resolves information on time position and frequency component. It is suited for non-stationary signals. Wavelet transforms [29, 30] work with dilation and compression of a special window function. It is applicable to non-stationary signals. Wigner-Ville distributions [28] give a time-frequency dependent energy density spectrum of the input signal. It generates new linkages of frequency components. It is suited for non-stationary data. Hilbert-Huang

transforms [31, 32] decompose the input signal into intrinsic mode functions. The input is represented as energy density over time. As well, it is applicable to non-stationary signals. Principal component analysis [33, 34] transform the original input into a new representation of uncorrelated features. Fisher linear discriminants separate the projection of input data in least square sense. Gaussian mixture models fuse the information in the input data in a probabilistic model [35]. The set up parameters are not easily definable. Logistic regression depicts the model fitting the best connection between input and output data [36, 37]. Statistical pattern recognition compares a given input data to a defined normal signal [38, 39]. It is only applicable for approximately normal distributions. Particle filters are a Bayesian approach on state estimation [40]. The input data are sampled to deduce a probability distribution function. It is applicable to non-stationary signals if the system dynamics are analytically defined. A high computational load is reported. Kalman filters are an other Bayesian approach on state estimation with covariance minimization [41, 42]. They are only applicable to linear systems and Gaussian noise. Again, a system model and a state model have to be defined. Feature map pattern matching reduces the feature space of the input data to a lower dimensional space [43, 44]. The orientation of the input space is maintained. No standard algorithm defining the map is given.

Bayesian networks give the dependencies of variables in an input signal [45, 46]. It is useful to reduce the number of parameters to describe a signal structure. The learning phase is rather complex and costly and expert knowledge about the modeled structure is needed. Neural Networks create a model of the relations between the input and output data and are able to detect patterns in a data set [47, 48]. The set up resembles to biological neuronal networks which is adaptable to unknown problem set ups, but no standard procedure for the network development can be given. The computational load is high. Fuzzy logic can offer robust and fault tolerant models from incomplete input data [49, 50]. Support Vector machines map the input data into a high dimensional vector space until data are separable [51, 52]. This classification method achieves good decision accuracy because of the maximal margin between the separating hyper plane and the nearest data point of a different class. The Hidden Markov model is a statistical model of a Markov processes representing the system [53, 54]. To generate an accurate model, a large amount of data is needed.

Table 2.1: Common algorithms and application fields of fault diagnosis (cf. Lee [10])

Application field Method	Bearing	Gear	Shaft	Pump	Generator
Time domain analysis	[25]				
Fourier transform	[55, 56]	[57]	[58]	[59]	[60]
Short-time Fourier transform	[61]	[62, 63]		[64, 65, 66, 67, 68, 69, 70]	
Wigner Ville distribution			[71]		
Wavelets	[72]	[21, 73]	[74]	[75]	[76, 77, 78, 79]
Hilbert-Huang transform	[80]	[81, 82, 83]			
Principal component analysis	[84]			[85]	
Particle filter		[86]			
Kalman filter		[87, 88, 89]			[90]
Neural networks	[91, 92, 93]	[94, 95, 96, 97]	[64, 65, 66]	[98, 99, 100]	[101, 102, 103]
Auto regression [104]	[105, 106]	[107, 108]	[109, 110]		
Fuzzy logic	[111]	[112]	[67]	[113, 114]	[115, 116, 117]
Match matrix [118]					
Support vector machine	[119]	[120]	[68]		
Hidden Markov model	[121]				[122]
Stochastic model	[123, 124]				

Table 2.1: Common algorithms and application fields of fault diagnosis (cf. Lee [10])

Application field Method	Bearing	Gear	Shaft	Pump	Generator
Genetic algorithms	[125, 126]		[68]		[117]
Empirical mode decomposition	[127]	[128, 129, 81]	[130, 131]		
Analytical or numerical models			[132, 133]		
Petri nets					[134]
Instantaneous power spectrum	[135]				
Bispectrum	[136]				[137]
Autoregression-fuzzy hybrid model		[138]		[139]	
Energy index analysis		[140]			
Envelope analysis				[141]	
High resolution spectral analysis					[142, 143]
Expert systems					[144, 145]
Higher order statistics					[146]
Park's current vector pattern					[147]

2.2 Time-frequency-based strip rolling mill applications

Strip rolling mills are complex steel production plants. The very large and heavy equipment has to operate precisely to achieve a high-quality product meeting the specifications demanded by customers. Therefore, research in the area of strip rolling mills concentrates on topics related to production, focusing on bearing fault detection, gauge control, gear fault detection, and chatter detection or damping. So far, data-driven approaches in strip rolling mills have received little attention in literature. A considerable number of contri-

butions applies model-based (MB) analysis, for example plant control models [148, 149], looper models [150, 151], roll shifting strategies [152], reheating furnace control strategies [153], chatter detection/damping [154], material science, or finite element modeling of strip curvature [155]. An overview is given in Table 2.2.

Table 2.2: Overview on relevant condition monitoring and signal analysis research areas in strip rolling [19]

Application field	Goal / Aim	Analysis method	References
Bearing fault detection	Extending life time	MB, FFT	[156]
Chatter detection	Improvement of surface quality	MB, LMD, STFT, FFT	[154], [157], [158]
Gauge control	Improvement of quality	MB, DWT	[151], [159], [160], [161]
Gear fault detection	Extending life time	MB, DWT	[162], [163], [164], [165], [166], [167]
Strip travel	Improvement of surface and flatness quality	MB, DWT, EMD	[168], [169], [170], [171], [170], [172], [173]

A variety of analysis methods applied to strip travel can be found in the literature. A detailed application to a hot strip mill process is given by Peng et al. [159]. The authors propose a data-based approach for online identification of the variables responsible for specific quality faults (hydraulic gap control, cooling valve control, bending force control). In this non-linear application, a total kernel projection to latent structures (T-KPLS) model with radial basis kernel function is used. The model is built with data from preceding periods. Process and quality data are evaluated. The result is presented in a contribution rate plot showing the fault responsible variables. This plot shows the sensitivity of specific faults to the examined variables. The three above mentioned examples of quality related faults are successfully identified by this approach.

Wang et al. [158] evaluate the use of local mean decomposition applied to low-speed helical gearboxes in matters of surveillance and diagnosis. The authors apply the instantaneous time-frequency spectrum resulting from local mean decomposition (LMD) to vibration signals of a gearbox of a finishing rolling mill to detect gear tooth damages. The authors propose a parameter to asses the severity of faults. This parameter was stated to be sensitive only for the wear of monitored components and not to be affected by

other influences such as load and speed changes. The application to practical data shows the efficiency and reliability of the applied approach in fault detection. The approach outperforms kurtosis, root-mean-square and peak-to-peak values.

Serido et al. [163] present a residual-based fault detection approach for rolling mills. As model architecture, a genetic box-cox model for linear components combined with a Takagi-Sugeno fuzzy model for non-linear components is proposed. The used data-driven soft computing techniques transform the original signals into this model space. According to the authors, no pre-developed analytical respectively physical fault model is used. For fault detection, the model-based calculated residuals are analyzed online with statistical techniques. The authors compare the performance of their approach to three state-of-the-art methods, namely principal component analysis, ARMA, and one-class SVM. Here, performance means the relation of fault detection (true positive detection of fault) and false alarm rate (false positive detection of fault). The author's approach achieves detection rates from 65% at slight fault symptoms up to 90% at strong fault symptoms with a false alarm rate of 10%. The tested fault symptoms are simulated. The total amount of used data is not given.

Li et al. [164] present an improved stochastic resonance approach, which is a method of using noise to enhance signals. This approach can be applied if the system has two or more unstable equilibrium positions in resonance phenomena. The noise can be used to switch the system between the equilibrium positions. Weak fault features are extracted from background noise to improve the detection of gearbox faults in a rolling mill. The used multi-stable model increases the signal to noise ratio and analyses the influence of resonance effects. The approach is tested on real data from a rolling mill. Two specific gear faults are detected successfully. Shortly later, the authors propose a modification of their method, the adaptive stochastic resonance with an additional sliding window technique [165]. In this approach, the window width can be adapted to the fault. The authors report an increased signal-to-noise ratio in simulation experiments and in practical application. The practical value of both techniques is stated. For weak signals, the adaptive technique performs better than the standard stochastic resonance approach.

Hui et al. [160] present a data-driven online algorithm for automatic gauge control. In steel production, gauge control determines the strip thickness and therefore has a major influence on strip travel and product quality. The authors combine multiple least square SVMs to reduce calculation time. The sample data are divided into different groups by subtractive clustering. Single least square SVMs are performed on each group. The sum of the weighted outputs of all SVMs gives the predicted thickness. The procedure was tested on a real hot strip mill. A total number of 300 data sets consisting of rolling force, roll gap setting value, entry thickness, rolling speed, and entry temperature are used. The

authors state the algorithm's performance is better than the ones of back propagation neural networks (NN) and single SVM.

Sanfilippo et al. [171] present an internet-based data reporting system for online monitoring in hot strip mills. The graphical reporting system considers the whole plant, starting at the furnaces and ending at the coil transfer. Here, cobble diagnosis is of special interest. Cobble is identified after the plant has stopped. Possible causes for cobble are retrieved from the data base and listed. A graphical representation of relevant process variables is displayed to machine operators. The distinction between good and bad conditions of each plant's element is based on failure mode and effect analysis (FMEA). This expert knowledge system is implemented using only commercial software elements. Details of the used mathematical models are not given.

Nandan et al. deal with a multi-objective optimization scheme and its application to a hot strip mill [169]. Opening with a detailed illustration of the scope of roll shifting the authors use distance-based Pareto genetic algorithms (DPGA) and strength Pareto evolutionary algorithms (SPEA) to optimize roll shifting owning to the scope of flatness and crown. The presented model is able to assess the hot rolling practice, but has not been implemented online. The study gives a quantitative overview of the connectivity of surface flatness, rolling schedule and roll shifting.

Arinton et al. investigate the use of artificial NN to handle the non-linear problems in a cold rolling tandem mill [174]. They present a multilayered dynamic high-order neural network based model of inter strip tension. Their approach is applied offline for modeling and residual-based fault detection to real data of a tandem mill. To increase robustness in fault detection, they propose to obtain the variable threshold from the confidence interval. The success of this kind of fault detection depends on the model's accuracy. Assuming a model with accurate estimation of the system's output the proposed NN approach is robust and useful in practice.

Debon et al. give a comparative overview of statistical models for binary data using receiver operating characteristic (ROC) curves [175]. They examine the use of this technique to visualize, organize and select classifiers based on their performance. The aim is to identify an optimal model to predict the probability of defective steel coils. Compared are generalized linear models and classification and regression trees on the basis of short time histories of temperature and velocity of a typical galvanizing bath. A generalized additive model was used to confirm the linear relationship. They state classification and regression trees as useful in steel coil quality prediction for practical needs.

Zhang and Liu propose a cascade predictive control strategy for hydraulic automatic

gauge control (AGC) of hot rolling mills based on data-driven control theory [176]. Their approach includes a secondary loop control system supervising the main AGC loop control system to overcome the problems of inaccurate indirect measurements and time-delayed direct measurements. Secondary loop control is a PID controller. Due to the use of data-driven control theory the model identification process can be avoided. The authors claim that their method is able based on simulated results to improve the control precision and to reject disturbances.

2.3 Strip travel applications of selected time-frequency-based analysis methods

The suitability of the following time-frequency-based analysis methods: STFT, WT, WVD, and HHT for industrial applications is emphasized by Lee et al. [10]. Disregarding that the resolution of STFT and WVD is only scalable in different runs, the disadvantages mentioned by Lee et al. are of minor importance to the diagnosis task. Based on their widespread application to fault detection in other areas [8], these methods have been checked for their use for fault diagnosis in hot strip rolling mills. Only a small number of applications to strip travel faults in rolling mills can be found in the literature.

Short-Time Fourier Transform

An application of STFT to chatter detection is given by Garcia et al. [157]. Chatter is a specific vibration in a rolling stand that affects surface quality and leads to thickness variation. The authors consider chatter-related features in the frequency band of 100-300 Hz. The used mathematical methods are spectrogram analysis and Fourier transform. A nonlinear dimensionality reduction with isometric feature mapping (ISOMAP) is performed on the spectrogram of STFT results. Parallel feature vectors containing the information of signal strength for predefined frequencies are extracted from FFT results. Both datasets are combined in a general regression neural network (GRNN). The result is given in a graphic showing characteristic zones related to chatter and non-chatter rolling to the machine operators.

Wang et al. [177] propose an autocorrelation-threshold-based method to extract the periodical components of vibration signals associated with mill chattering in strip rolling mills leading to chatter marks. Chatter marks are unwanted fringes on the strip surface resulting form mill chatter or insufficient lubrication. Chatter features are hidden in noise. Since the occurrence frequency is not constant, the application of a band pass filter to

reduce the noise is not useful. Autocorrelation is able to highlight the frequencies occurring periodically. On these enhanced frequency components FFT is applied to identify the chatter. The results still contain unwanted components.

The contribution of Garcia is the only recent application of STFT to strip rolling mills that could be found. The contribution of Wang uses FFT after an autocorrelation generated short time window of the original signal. Both contributions aim on chatter detection.

Continuous Wavelet Transform

Applications of CWT to strip rolling mills could not to be found.

Discrete Wavelet Transform

Chen et al. [172] give a DWT application on roll eccentricity that causes thickness variations in hot strip mills. Product quality and strip travel are possibly affected. The authors use a multi-resolution wavelet method that combines mallat algorithm (FWT) with FFT and inverse FFT (IFFT) to overcome the aliasing caused by the cutoff frequencies of the wavelets. Wavelet functions are not designed to minimize the aliasing. Both FFT and IFFT are used in wavelet decomposition and reconstruction procedures to set those components to zero, that lay outside of the aimed frequency band. The approach is tested on a six-stand four-roll-continuous hot strip finishing mill. For the DWT, a sym5 wavelet is used with six decomposition levels. The algorithm with increased calculation time was tested on the machine. The in-time compensation in gauge control was able to reduce thickness variation from $\pm 40\,\mu$m to $\pm 15\,\mu$m.

Li and Dong [161] present a wavelet and neural network-based fault diagnosis approach for the hydraulic automatic gauge control of a strip rolling mill. This moving average model consists of a three layer forward network. It is supposed to improve rolling force forecasting in real time. Diagnosis is achieved using wavelet transform of the residuals of predicted rolling force (setting signal) and actual rolling force (sensor signal). The Haar or DB1 wavelet transform is able to detect the position in time of the occurrence of a fault. The maximum wavelet coefficient is found at the step in the residual signal. The degree of fault is indicated by the coefficient value of the DWT. A NN builds the model of the system. Actual field data are used to establish the model structure. The type of fault is identified by comparison with data from this model of the system. Additionally, the wavelet transform de-noises the signal. The method is tested successfully offline on recorded data of an inner-leak servo fault.

In addition, examples of extended wavelet transform algorithms exist. Using an approach of Mallat and Zhong [173], Lesecq et al. [162] treat vibrations in the production line that

may affect surface roughness and quality. According to the authors, these vibrations can be induced by main drives, badly parameterized controllers, frequency converters, torque changes, or defect components. A time-invariant stationary wavelet transform (SWT) is used to detect specific vibration faults in a rolling plant's signals with fuzzy decision making. Real data of a roughing mill's twin drive are applied namely torque, current, and speed of upper and lower motor. First, an SWT is performed on the signal, followed by soft thresh-holding, fuzzification, and aggregation leading to a symptom. The amount of testing samples was low (5). Only torsional vibration faults were detected.

Yuan et al. [166] present a fault diagnosis approach for a rolling mill's main drive gearbox. This is based on multi-wavelet sliding window neighboring coefficient denoising and optimal blind deconvolution. The sliding window technique tries to solve a shortcome of universal wavelet thresh-holding. Important fault features with only small signal components might otherwise be masked. The sliding window cuts out time slices with individual thresholds for weak signals. The relativity of conducted local threshold coefficients to their neighbors is used for further denoising. Blind deconvolution is used to sharpen and isolate the implied fault features to be more easily recognized by machine operators. The approach is applied to two real gearbox fault data sets of a finishing mill. The authors point out the practical possibility to detect multi-faults characteristics and to avoid missing weak features. Optimal blind deconvolution is good in detecting transient information but struggles from heavy disturbances.

The number of recent contributions applying DWT in the field of strip rolling mills shows some research work in this topic. The number of contributions is considerably higher than the number of applications of the other methods discussed. The success of the application of DWT in production is shown by the contribution of Chen et al. [172]. Additionally, the amount of applications of DWT in different fields of fault detection indicates that this topic will be important in future.

Wigner-Ville Distribution

Applications of WVD concentrate on periodic signals and cannot be found for strip rolling mills.

Empirical Mode Decomposition

An application of EMD is given by Liu et al. [170], focusing on the remote fault diagnosis of heavy mills. The authors present a four layer system consisting of a data hunting layer where the sensor signals get transformed, a knowledge management layer where the data are stored, an application layer where signal processing is performed, and an user interface.

The pattern recognition is done with a SVM, the fault location with EMD-HHT, and the remaining useful lifetime prediction is done with SVM regression. Amplitude, frequency, and kurtosis of vibration signals measured on seven key points are used as characteristic values. The authors do not give any application or simulation results.

Only one contribution of EMD to strip rolling mills could be found. The method is not yet widely spread.

2.4 Summary

The application areas of the large variety of fault diagnosis methods focus on periodic signals mainly in rotating devices. In the field of signal-based analysis of strip rolling mills only a small number of published results can be found. The focus of researchers lies on rotating machinery and concerns mostly periodic or quasi-periodic signals. Overall, applications to non-stationary signals are rare. Introducing new concepts to production is difficult, since online experiments can lead to costly mistakes. Therefore, all but one tests found in the literature are made offline.

The review on literature dealing with the five selected methods shows that only three of them have been applied to strip travel in rolling mills: STFT, DWT, and EMD. The application of STFT concentrates on chatter detection with mainly periodical signal components, the application of EMD focuses on vibration signals. The applications of DWT spread over a wider field of process deviations containing periodical as well as non-stationary signals.

The question whether or not these promising techniques - which were useful in a wide field of industrial processes - can be exploited for the diagnosis task on strip travel in hot strip rolling mills has to be answered. Therefore, it can be concluded that further research is necessary. This work applies several signal-based diagnosis methods in the field of strip travel faults in rolling mills. The suitability of the methods for fault detection and fault identification is evaluated, and a prognosis is deduced.

3 Introduction to the application site

Steel, a metal alloy with iron as principal component, can be given a defined shape in forming processes. The forming process is a manufacturing process aiming to a plastic deformation of a material into a defined shape [178]. One of these forming processes is rolling. Forming enables the production of special grades and properties that cannot be achieved by casting.

To roll a strip, a steel slab is needed as primary material. The production of a steel slab is performed via ingot casting or continuous casting. Beside billets, blooms, and monocasts, steel slabs are the first shaping product of steel after liquid phase. The rolled steel strips are categorized according to the thickness. Steel strips thicker than 3mm are called heavy plates and strips thinner are called thin sheet steel. In the basic concept, a two high stand, the metal is passed through a pair of rolls to reduce the thickness and to enlarge the length. A uniform surface, width, and thickness is aimed for. The metal is heated to allow the forming process. Depending on the temperature, the process is called hot or cold rolling. In case of hot rolling, the rolling temperature is above the recrystallization temperature. Hot strip describes a warm and flat rolled forming product. Various grades of steel with respect to mechanical properties and surface demands lead to differing production conditions. Consecutively rolled strips may demand varying parameters in the control circuit. Hot strip is the basis for diverse application fields, for instance in mechanical engineering, shipbuilding, automotive industry, bridge construction, and container construction. Applications of direct processed hot strips are for example tubes and tanks. Depending on the field of application, the steel strip is subsequently processed in other aggregates for fine tuning or surface finishing. Depending on the steel type, additional coatings are available. Thin sheets can be enameled, galvanized, nickel-plated, painted, tinned, or plastic coated [7]. The hot rolled product is delivered in coils. The finished strips show a manifold range of quality grades, tensile strengths, and bending strengths. This section gives an introduction to flat rolling in hot strip mills and selected strip rolling problems. The content of this chapter is based on contributions already published [19, 179, 180, 181].

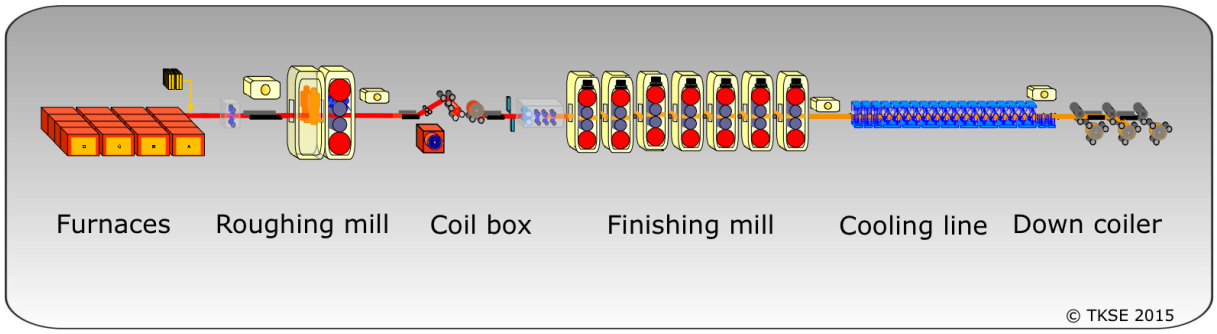


Figure 3.1: Hot strip rolling mill of TKSE in Bochum [7]

3.1 Fundamentals of a seven stand hot strip mill

The approach proposed in this thesis is build on real data originating form the hot strip mill of ThyssenKrupp Steel Europe AG (TKSE) in Bochum [7]. The hot strip rolling mill in Bochum is rolling semi-continuously. Primarily, it consists of four reheating furnaces and two pit furnaces, a roughing mill, a coil box, a finishing mill, a cooling line, and down coilers. An overview is given in Figure 3.1.

The furnace system is composed of three walking-beam furnaces and one pusher type furnace (see Fig. 3.2). These furnaces are natural gas-fired continuous furnaces, in which the slabs lay transverse to the transfer direction. In a pusher type furnace, the slabs are pushed through a hot-cooled supporting tube system. In a walking-beam furnace, the slabs are moved by a horizontally and vertically moving walking beam conveyor. By lifting the exposed slabs, the slab's surface is less stressed during the reheating process. Certain special grades, for example non-ferrous alloyed steels, are soaked to rolling temperature in a pit furnace. The rolling temperature depends on the material to be processed. Temperatures between 1200°C and 1300°C are common. Special grades may differ in temperature, e.g. titanium is rolled at about 800°C.

Discharged from a furnace, a reheated slab is transferred to the roughing mill compound starting with the first descaler (see Fig. 3.3). The iron-oxide is removed with 125 bar hydraulic thrust from the slab surface to prevent scrap marks. The consecutive guiding side system centers the slabs for threading into the edger. The edger is a vertical stand sizing the strip width. A four-high reversing stand is following directly. It consists of four horizontal rolls. The two thicker backup rolls reduce the deflection of the thinner work rolls. This reversing stand is passed five, seven, nine, or eleven times reducing the entry slab thickness from 150-260 mm to 35 mm.

After the roughing mill, the resulting transfer bar passes a roller table containing a coil box, a cropping shear, and a second mechanical descaler (see Fig. 3.4). In coil box mode,

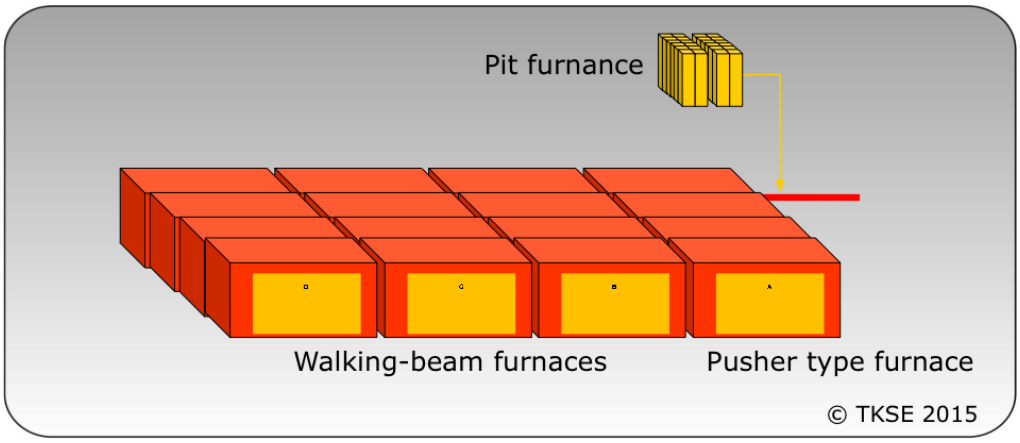


Figure 3.2: Furnace compound [7]

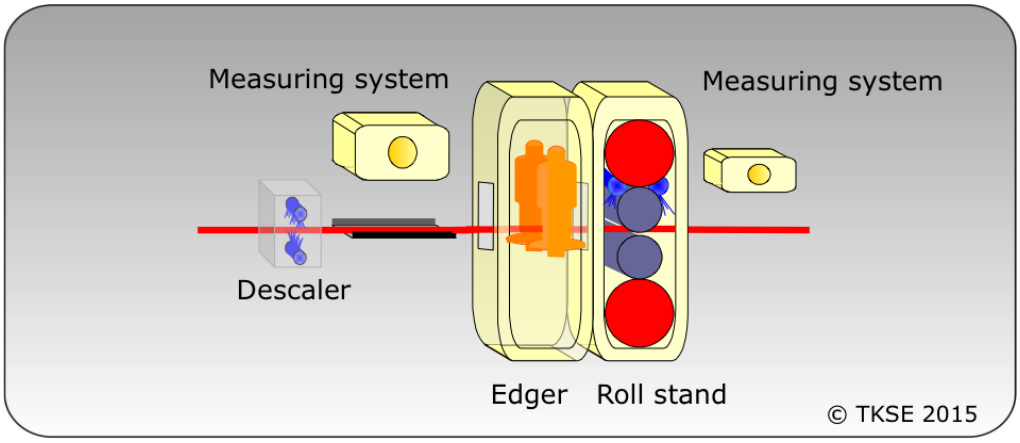


Figure 3.3: Roughing mill area [7]

the transfer-bar is coiled to prevent unwanted thermal effects, like irregular cooling across the length. The transfer-bar is threaded bottom up into the finishing mill. The homogeneous temperature allows a homogeneous roll force. At this point, materials reheated or soaked in a coil box furnace can be re introduced into to the process line. The cropping shear straightens the head of the transfer bar to reduce alining difficulties during the initial pass section. Likewise, the tail is cropped to avoid deformations as e.g. fish tails. The cropped bar is descaled a second time before threading into finishing mill.

The finishing mill consist of seven continuous rolling four-high stands (see Fig. 3.5). Each stand contains a pair of work rolls and backup rolls. The diameter of the backup rolls is higher than that of the work rolls. This gives a higher mechanical stability and reduces bending of work rolls. These rolls have to be exchanged regularly because of the abrasion. To ease the maintenance, each roll is mounted by a chock. Between the stands, a looper with adaptive angle supplies a nearly constant mass flow. The roll gap

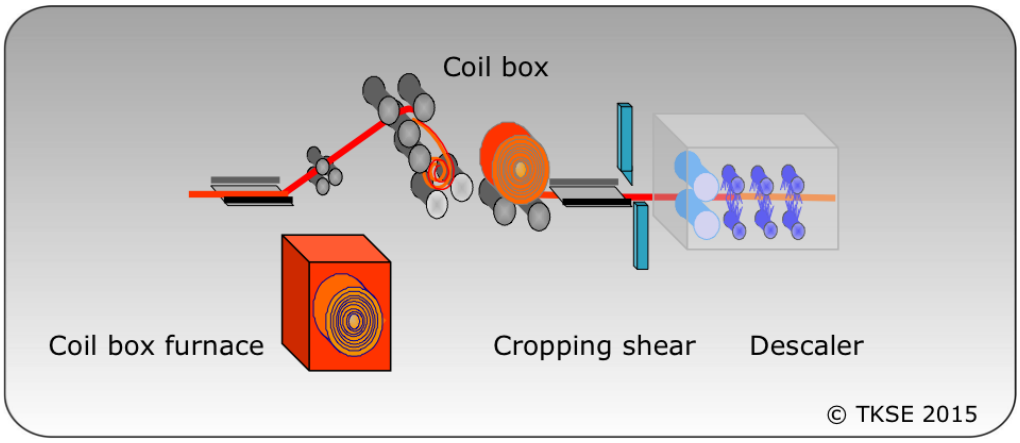


Figure 3.4: Roller table [7]

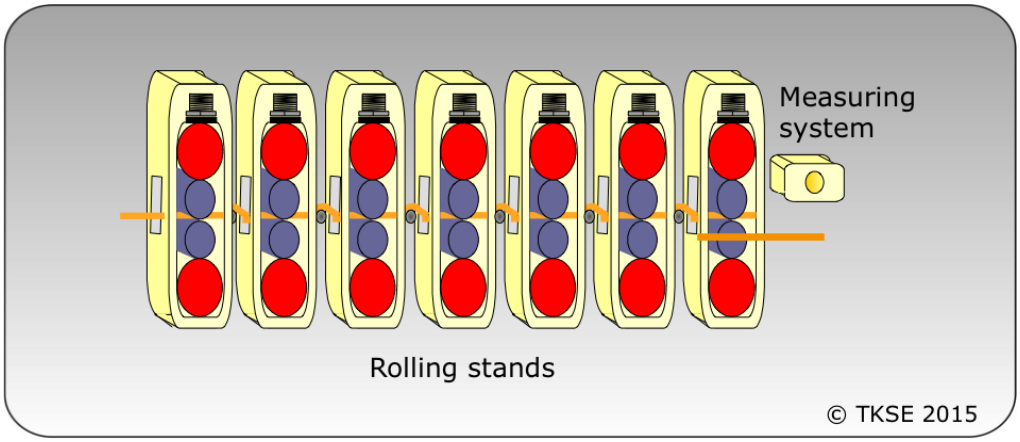


Figure 3.5: Seven-stand finishing mill [7]

is adjusted by a hydraulic mechanism. The continuously variable crown in stands 3, 4, and 5 gives the possibility to influence profile and flatness to grant a constant course. The roll force depends on the amount of forming to be made and on the deformability behavior of the material. The target thickness of the roll gap is adjusted by a hydraulic screw down. The work rolls are joint by a pinion gear unit to two coupled DC motors. This synchromesh gearbox drives work rolls with the identical speed. The backup rolls adapt through friction. The roll force is measured with load cells, located underneath the lower backup roll's chocks. Using magneto-elastic effects, up to 40 MN roll force can be captured. In practice, roll forces up to 30 MN are used. Typical physical dimensions of a finished strip lay between 1.4–20 mm exit strip thickness, 600–1630 mm width and up to 1 km length.

Having passed the finishing mill, the finished strip is cooled down with respect to the client specific material requirements. The computer-controlled cooling line reduces the temperature from finishing to coiling values. Diverse cooling-down strategies are used to meet the cooling rate, which has the desired effect on material structure. Subsequently,

the strip is coiled down. In Figure 3.6, the cooling line and the down coiler are illustrated.

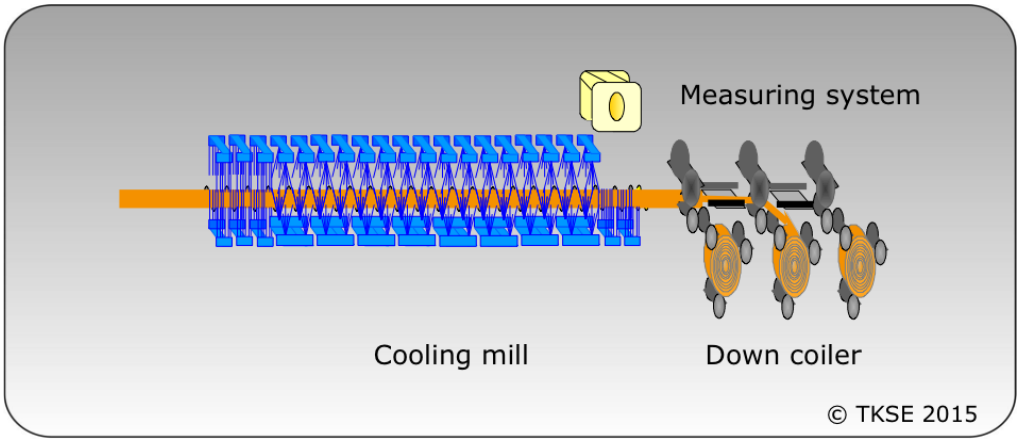


Figure 3.6: Cooling line and down coiler [7]

3.2 Technical fundamentals of rolling

The forming process realized here is flat rolling. It is used to reduce the cross-sectional area of a semi-finished product, to stretch it, and to define its material characteristics. The material structure, the surface properties, the profile, and other features are defined by the process temperature, the type of treatment, and the type of cooling.

Change in the shape of a metal is caused by displacement of the atomic lattice structure. Elastic and plastic forming are distinguished. In the elastic range, the displacement of atoms is so small that they return to their original lattice position after the removal of stress. This is not the case for plastic forming. Here, the lattice structure is permanently modified. Plastic forming of rolled material is achieved when the so called deformation resistance is overcome. This empirical value is *inter alia* dependent on the material properties and the temperature. This is used in hot strip rolling where a higher temperature is applied to reduce the required force. Since a preheated metal is more ductile, meaning its tension values decrease, strong deformation is possible without loss in material cohesion at very high temperatures [182].

The amount of deformation φ_h is usually given by the logarithm of the relative strain h_1/h_0 [183]

$$\varphi_h = \ln \frac{h_1}{h_0}. \tag{3.1}$$

The relative deformation is

$$\varepsilon_h = \frac{h_0 - h_1}{h_0}. \tag{3.2}$$

In the considered finishing mill, relative deformations ε_h between 0.21 and 0.29 are typically achieved. Flat rolling is supposed to produce a plastic stretching of the material by thickness reduction. This lateral expansion Δb is highly dependent on the reduction in thickness Δh

$$\Delta h = h_0 - h_1. \tag{3.3}$$

Especially in the range of the finishing train of a hot strip mill, the ratio b/h is as large that the absolute lateral extension amounts to only a few millimeters.

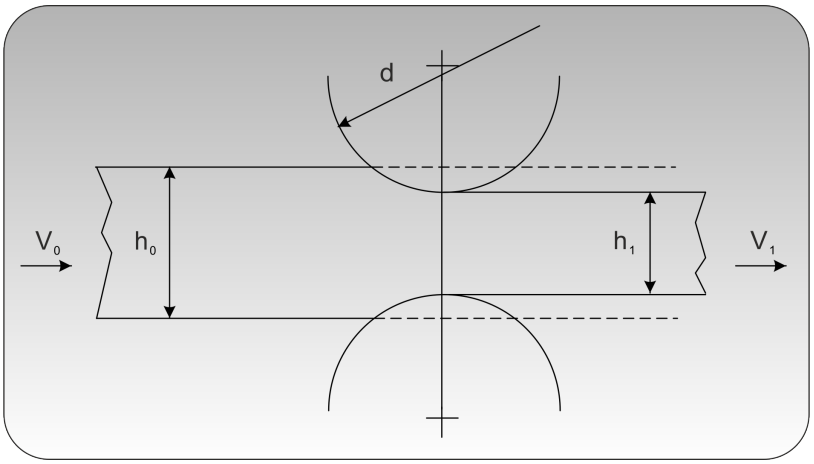


Figure 3.7: Simplified idea of flat rolling modified according to [183]

Hollow-ground cylindrical work rolls are used in the considered finishing mill. Figure 3.7 visualizes the simplified forming geometry of an idealized roll. At loaded contact of the rolls with the rolled material an elastic flattening of the rolls is determined. With sufficient mechanical stability, this roll flattening can be negligible in first approximation. The roll flattening is considered in the following formulas. The meaning of the parameters can be taken from Figure 3.8. The changed contact length is

$$l'_d = \sqrt{r' \cdot \Delta h}. \tag{3.4}$$

The contact surface A'_d between roll and rolled material depends on the input width b_0

$$A'_d = l'_d \cdot b_0. \tag{3.5}$$

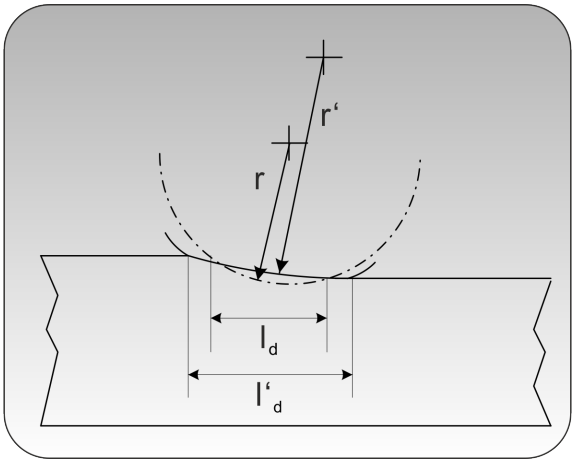


Figure 3.8: Simplified idea of flattened roll modified according to [183]

Starting from an unencumbered contact of the work rolls, the screw down of the roll adjustment generates a load that bends the components of a roll stand in the elastic range. The floating Morgoil®chucks bring an additional elasticity to the compound. Also, the specific plastic-elastic behavior of the rolled material influences the roll gap. These and further details have to be considered in any analytic or physical model of a rolling process. In this work, no model of the rolling compound is used for this research. Instead a signal-based analysis approach was chosen.

3.3 Deviations in strip travel

Strip travel describes the way heated rolling material passes the rolling mill. Ideally, the rolling material passes all aggregates in center position, without any surface defects, within the thickness tolerance, and with optimal flatness and profile. In practice, not all of these goals are achieved in each run. Some of the deviations lead to repairable defects on the strip, for example indentations and protuberances [184]. Other defects are not repairable, for example shells. Depending on the required quality and the type of fault, the rolled material has to be refurbished, devalued or wasted. Strip travel problems appear as wavy edges, profile out of tolerance, and many more. In the presented contribution, two specific strip travel deviations are discussed that affect product quality and cause downtimes: cobbles and shearing tails.

In this context, a cobble is a severe fault in the rolling process. It can appear in the area

Figure 3.9: Illustration of a deviation in strip travel called cobble [7]

of the roughing mill, of the finishing mill, and of the down coiler. The events appearing in the finishing mill will be discussed here. In case of a cobble, the rolling process is run as usual, but suddenly the strip bows up between the stands, high enough to loose the contact to the looper. The strip tension and therefore the mass flow in the roll gap is no longer under control. The rolled strip does no longer pass through the rolls. Instead, it becomes twisted and folds to many slings that pile up in between the rolling stands, compare Figure 3.9. The rolling process can not be continued and the cobble has to be removed manually. In general, the material has to be sliced to be removed. This is a time-consuming process, leading to breaks in the production process. The break in production will last at least half an hour or up to a whole day ore even longer, if secondary demolitions occur on the aggregates. Depending on the grades, a long term break may damage the slabs remaining in the oven. Certain grades cannot persist for long time in the oven. Additionally, certain grades must not be heated twice, otherwise the aimed structure will not be achievable. To avoid those costly breaks and to enable a continuous production, it is highly desired to prevent this kind of fault.

Another deviation in strip travel is called shearing-tails, compare Figure 4.1. This fault occurs at the end of a strip as the name tells. Again, the rolling process is run as usual. But at the end of the strip, in the tail region, the strip breaks. The damaged strip parts run through the rolls, get folded and leave marks on the rolls. Once this fault is noticed, the damaged rolls are removed and exchanged immediately. Exchanging the rolls usually takes less than half an hour. Afterwards, the control loops will be adapted via a calibration process called “facing” and the rolling process can be restarted. If a shearing tail is not noticed immediately, the marks on the roll will leave scratches, grooves, and gouges on the surface of the consecutive strips and affect their surface quality. Broken parts of the

Figure 3.10: Illustration of a deviation in strip travel called shearing tail [7]

sheared strip may remain as obstacles in the roll gap have to be cut off. These obstacles would cause further demolitions to the aggregate and might lead to a cobble like scenario. Therefore, it is desirable to prevent or at least recognize such an event.

The time-signal of a single strip without known deviations is exemplarily shown in Figure 3.11. The graph on the top visualizes the normalized roll force and the bottom graph illustrates the angle movement of a looper.

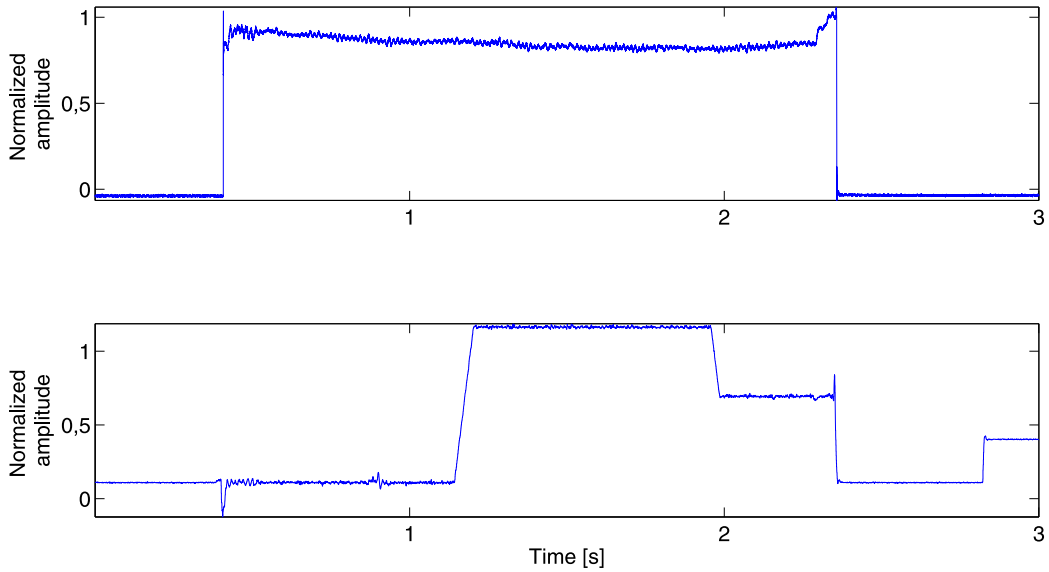


Figure 3.11: Example for a time signal of rollforce and looper angle

4 Development of a new signal processing method for fault diagnosis

In Chapter 2, the publications presenting signal-based analysis method for strip travel in hot strip mills are disussed. The lack of research in the field of fault diagnosis concerning cobbles and shearing tails is pointed out. The commercial need of the analysis of these deviations in strip travel is summarized in Chapter 3. In the context of this work, signal-based processing methods for fault diagnosis, focusing on these deviations in strip travel of hot strip mills were developed. Selected time-frequency analysis methods weretailored and tested for their aptitude in fault detection, fault identification, and fault prognosis. The content of this chapter is based on contributions already published [19, 179, 180, 181].

4.1 Definition of system states

Considering the rolling process of a hot strip rolling mill, two common system states catch the eye. First, a regular rolling process without known deviations. This will further be referred to as State 1. Second, an idle phase between to slabs, since hot strip mills are not rolling continuously in contrast to cold rolling mills. This idle phase will further be referred to as State 2. For this approach, these two states are defined as fault-free system states of the finishing mill. State 1 with strip in stand, State 2 without strip in stand. These two states are taken into account because of their frequent occurrence.

Considering the deviating system states of the finishing mill, two different fault types will be analyzed in this approach: cobble and shearing tail. The fault symptoms and their possible consequences are described in Chapter 3. These two faulty states are taken into account because of the severity of the damages emanating from them.

Figure 4.1 gives an imprint of the selected system states. The top-left image presents State 1, a fault-free case. The strip passes through the finishing mill without any disturbances. The top-right image visualizes State 2. Equally, it is fault-free. But in difference to State 1, it is a case without a strip in the stand. The bottom-left image illustrates

State 3. It is a fault case, namely a cobble. The strip is wound up between the stands.

The bottom-right image shows an other faulty case, in this case State 4. The crack in the strip and a rupture are visible.

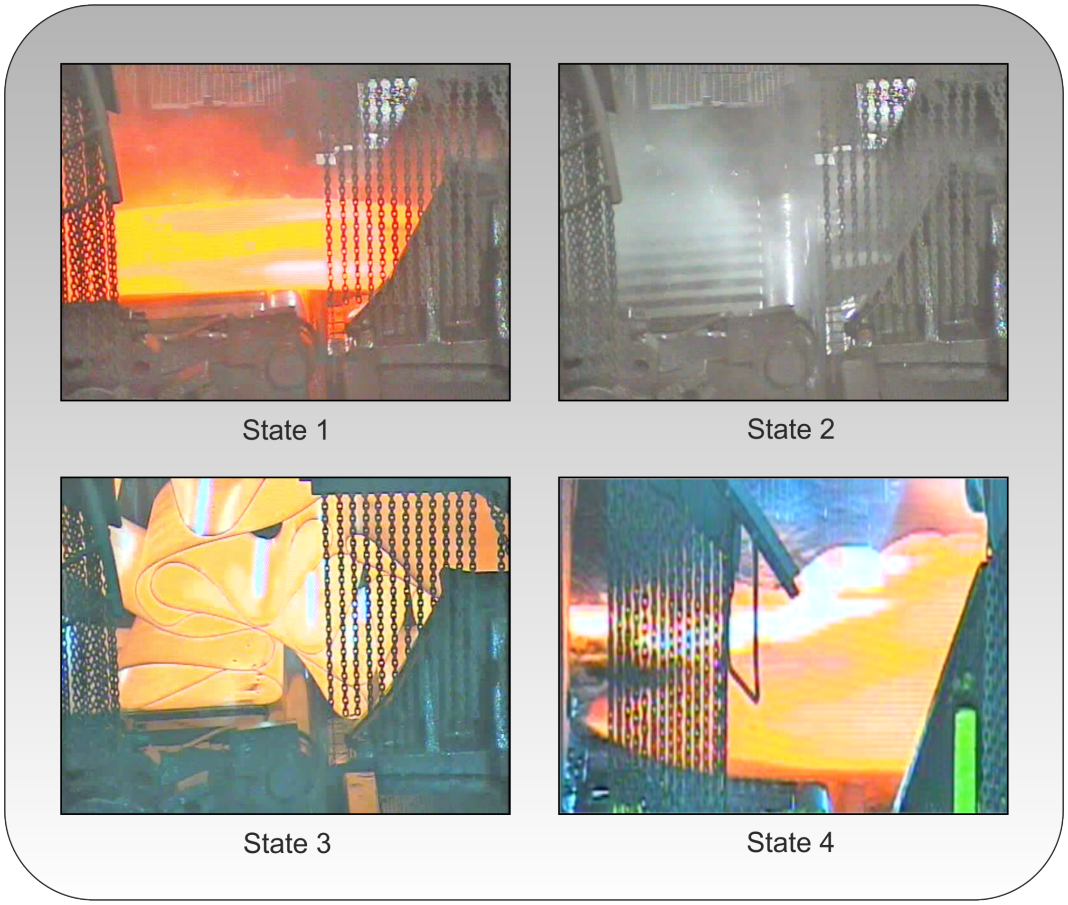


Figure 4.1: Illustration of the effects described by the four system states [180]
State 1: rolling; State 2: idle; State 3: cobble; State 4: shearing tail

4.2 Measurement selection

During production, a large number of sensor, control, and event data are captured and stored in specific data bases. Additionally, parameters originating from the control models are stored. Avoiding the use of models for control, only the measured signals will be taken into account. Still, the number of possible input data is huge and the origin is various. For example, the following values are measured: temperature, load, velocity, lubrication, dew-point, roughness, thickness, profile, center position, edge quality, and grain size. To start an analysis of the product line condition, appropriate signals have to be selected. In coordination with experienced machine operators, measurement data without known

influence on strip travel are omitted. This reduces the number of possible input signals from over 2000 signals stored on the iba® file server to 144 possible input signals.

Considering the two faults, cobble and shearing tail, the roll gap has great influence. Based on previous work [185] the roll gap is considered as a crucial element in rolling. In [185] a MATLAB Simulink® model of a four-high stand was build and has shown that the behavior of a stand is mostly influenced by roll-force parameter. The measured roll force contains implicit information about the condition of the roll gap. For this reason, it is selected as input signal for the fault diagnosis of strip travel.

The roll force is measured by load cells underneath the chocks of the backup rolls. In each stand, two of such load cells are used, one on the drive site and another on the operator side. Due to the continuous change in control parameters during the rolling process, the absolute value of roll force is not considered, but the difference of operator and drive side. To measure the roll force, a Millmate® roll force system with a Millmate 400® controller is installed. The sensor uses the magneto-elastical effect. Under load conditions, the magnetic properties of material are influenced. This change is measured and an output signal is generated proportional to the applied force. Details on the basic concept of the sensor are given in [186]. For the data used in this approach, the sensor output signal is sampled with 1kHz and stored on an iba® file server. In Figure 4.2, a simplified four-high stand is shown and the position of the load cells in the bottom part of a four-high stand is indicated by two arrows. The cells are not visible when the rolls are mounted. Figure 4.3 shows an original load cell mounted in a roll stand.

Figure 4.2: Position of a load cell [7]

As shown in Chapter 2, only little research in the field of time-frequency-based fault diagnosis has been done so far. So far, the roll force has not been taken into account for time-frequency-based analysis. In the present approach, the roll force as an essential process parameter is considered as basis for the evaluation.

Figure 4.3: Mounted load cell (rectangular box in the middle)

4.3 Data set selection

Event-based data sets stored by the machine operators give trustworthy information about the system condition. They are completed by the information resulting from further analysis steps executed by machine experts after the events. The sample data sets used for the present approach are selected according to the information in the event-based data.

For the selection of the sample data sets for the faults, the events assigned to the two fault types appearing in a certain time period are taken into account. Only two constraints are applied. The first is, that the strip has been threaded successfully into stand number 1. This kind of threading problem is to be solved by an other control unit and will not be considered in this approach. The second constraint is that no cold tips are treated. The temperature is an important factor in rolling. If the temperature delta is too high, the control parameters for the process automation must be adapted. Sensor systems in the production line are meant to solve this problem. Therefore, cold tips are not considered in this approach. No further pre-selection in matters of material group, material width or material thickness has been done. The data sets for the regular rolling process without known deviations and for the idle phase between two slabs are selected from the data base avoiding the same facing period as for the fault data set.

The time stamp of the collected sample data sets is checked and adapted if aligned with the time stamp of the event-based data. The data basis for the present approach consists of 80 data sets, 20 for each system state. Following a procedure by Kohavi [187], the data base is broadened by a four fold cross-validation. The total number of data sets after cross-validation is 320, i.e. 80 sets per class.

4.4 Signal processing techniques

The time resolved output of a load cell is exemplarily visualized in Figure 4.4. The normalized roll force amplitude[1] is plotted over time. During the depicted time period of 18 minutes, six strips have been rolled. They are clearly separated by the idle time between them, where the rolling force is about zero. During the rolling period of strip number 4, the machine operator aborted the rolling process due to process deviations recognizable with expert knowledge. The operator stored corresponding event-based data on a server and mentioned an upcoming cobble. In the raw time signal, the upcoming process deviation is not visible. Advanced signal processing techniques have to be applied to extract the necessary information about the system state.

The proposed treatment of roll force signals is given in Figure 4.5. The measurements of the load cells are acquired as described above and stored on an iba® file server. According to the event-data logged by machine operators giving trustworthy information about the systems condition, the sample data sets are selected for each system state. After a certain preprocessing, the fault features are to be extracted. Here, advanced methods for feature extraction and classification are used, namely STFT, CWT, DWT, WVD, and EMD.

[1]For confidentiality reasons.

Hereafter, the suitability of the methods to extract hidden information from the signal is tested. Different diagnostic approaches have been evaluated for their suitability in fault detection, fault identification, and fault prognosis. The approach of the classification step is statistically validated via hypothesis test and McNemar's test.

The following subsections give the mathematical background and details on the application of the selected signal processing techniques. It is divided in feature extraction step, classification, and validation.

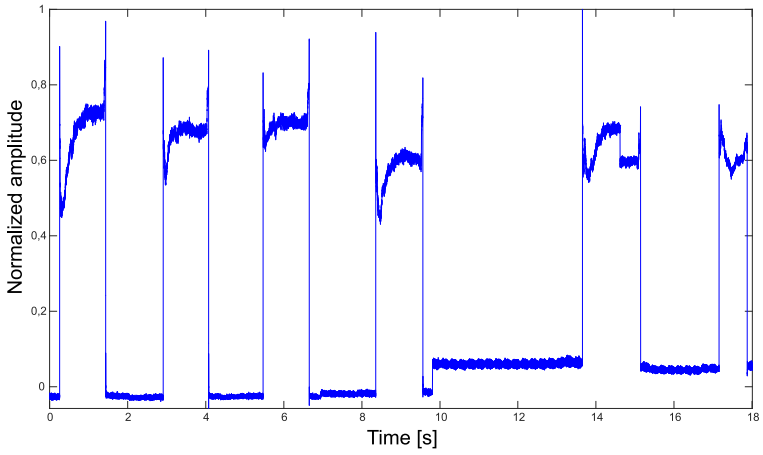


Figure 4.4: Time signal of a load cell [7]

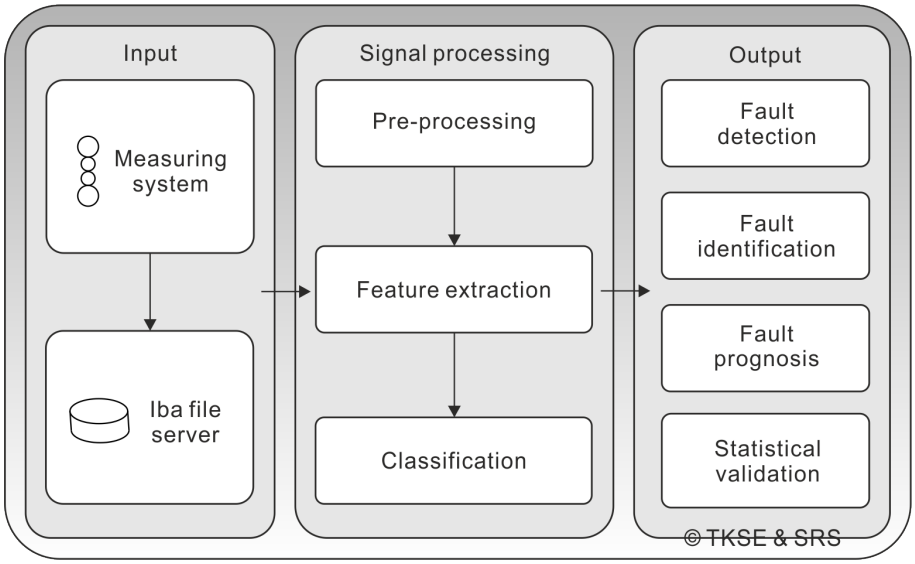


Figure 4.5: Proposed signal processing chain

4.4.1 Preprocessing

The input data are stored in a coded format on an iba® file server. To import the data into MATLAB®, a special interface called ibaFilesLite is used. This tool enables the decoding of the time-based data in MATLAB®. During the import sequence, a plausibility check is performed on the channel address and the data vector length. If the check is not passed, an alarm notice is set. The user has to confirm the alarm notice and has to validate the input channels before the import routine is continued. The imported data are decoded and stored in a MATLAB® table. After that, frequencies higher than 100 Hz are low-pass filtered. It is important for the classification task, to balance the length of input data vectors as an unbalance may influence the classification rate. Therefore, the filtered data are binned groupwise into pre-defined segments. Each group contains the same amount of sets for each system state. The length of a strip is individual, since it depends on the customer's order in weight, the individual rolling temperature, the individual lattice structure, and many other parameters. So, the length of a strip is not a possible input data length. In this approach, a time window of five seconds was set as input data length. All sample data set segments take a time window of five seconds.

The information on the system state is derived from the event-based data stored on a SQL data server depending on the arbitration of the operator. These data are used to define the class of each input segment, thus selecting one of the four machine states. The class of each input signal is passed on as a label vector. To coordinate the information, the label vector is concatenated as first column to the input signal. During the further advanced signal processing approaches, it is ensured that the label vector is not part of the input data.

4.4.2 Signal processing techniques for feature extraction

For feature extraction, five time-frequency-based data analysis methods are used. The selection of these methods is described in Chapter 2. This section will give a basic introduction to the methods and the individual adaption to the diagnosis task in hot strip mills.

Short Time Fourier Transform

Fourier transform (FT) is one of the oldest methods for signal analysis. It was presented by Joseph Fourier in 1822 and is defined as

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t} dt, \tag{4.1}$$

where t is the time and ω is the frequency parameter. It gives the spectrum of the time series signal $x(t)$. The FT is meant to show the frequency components of a signal. Due to its continuous nature and the necessary mathematical conditions, only periodic signal components are detected. To overcome this drawback, Gabor developed the STFT in 1946. The STFT results contain information about the temporal alteration of frequency. The transform slices the signal into short-time windows, which are treated as quasi-stationary, allowing to localize the frequency components with respect to time.

From a mathematical point of view, the FT is only valid for an indefinitely lasting signal. In standard FT as well as in STFT, the window function delimits a time slice and has a substantial effect on frequency resolution and apodization. The window width, however, has an additional effect on the results. The use of a small window achieves a good time resolution with a bad frequency resolution as trade-off and *vice versa* as given in Fig. 4.6. For discrete data, the discrete Fourier transform is applicable. The discrete formulation of the STFT is

$$X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]\, w[n-m]\, e^{-j\omega n}, \tag{4.2}$$

where $X(m, \omega)$ are the Fourier coefficients depending on time index m and frequency ω, $x[n]$ are the data points at time index n, and w is the window function; see for example [188].

The STFT is one of the most popular methods in practical applications. The representation of the results in the time-frequency plane or a spectrogram is easily understood. One problem in applications is to find a suitable window size.

Schlagner [189] shows that STFT is only suited for weakly non-stationary signals, since using a fixed window size yields the same time-resolution for all frequencies. Instead of STFT, the author proposes the wavelet transform (WT).

In the present approach, the STFT is executed with the MATLAB® on-board function *spectrogram*. An input signal contains 1000 sampling points. A Hamming window with a width of 256 samples is applied (compare Figure 4.7). The Hamming filter is used,

since it strikes a good balance between a low number of overshooting and a low loss of time resolution. It has been tested that exchanging the filter window doesn't lead to better transformation performance. The window is shifted with 250 overlapping sample points between adjoining segments. This is resulting in 125 different units characterizing the time position in the signal shape. The detail is responsible for the expected time behavior.

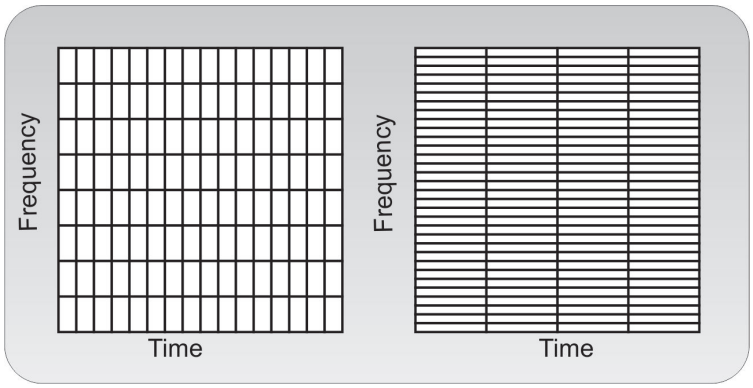


Figure 4.6: STFT resolution as a function of filter window width [19]

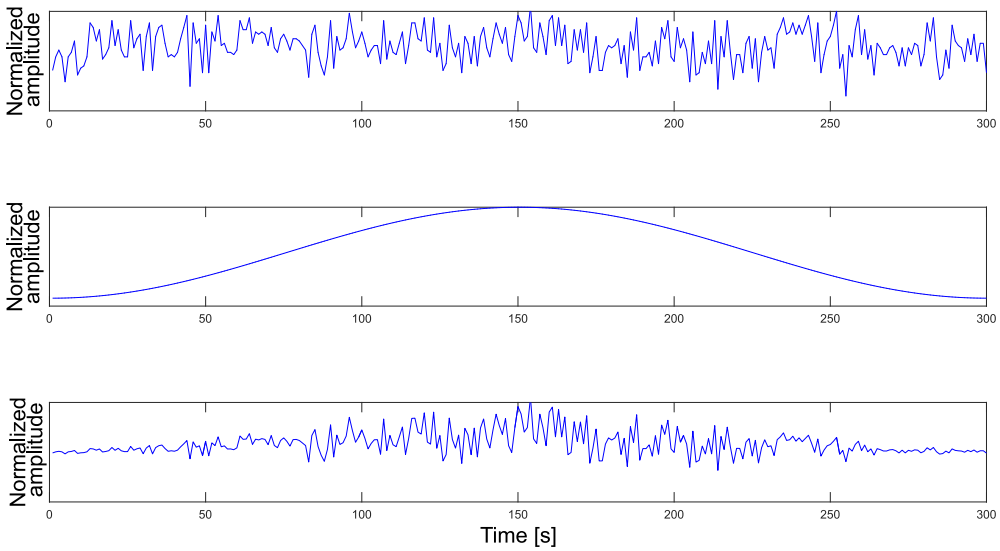


Figure 4.7: Application of a Hamming window to time signal; top: Raw time signal; Middle: Hamming window; bottom: Filtered Signal

Continuous Wavelet Transform

The WT is a signal processing tool used in different fields, such as speech analysis, image analysis, and data compression. The term wavelet was first used by Grossmann and Morlet in 1984 [190]. The name is deduced from "little wave" and is based on its wavy shape. A historical overview and mathematical basics can be found in [191], [192], [193]. This multi-scale analysis method is able to give a suitably related resolution in the time and the frequency domain, as depicted in Figure 4.8. The resolution problem can be resolved by adopting the window function. Typically, multiple shrunk and widened versions of the same window function are applied. These window functions are called wavelets. The amplitude of a wavelet starts and ends at zero, and its definite integral has to be zero. Therefore, the wavelet's envelope is often fading out peripherally. The shape and the characteristics of wavelets are diverse. Most common wavelets are Haar, Daubechies, Meyer, and Morlet wavelets.

Lin and Qu [194] give a detailed introduction to wavelet transforms with a focus on Morlet wavelets. The authors use the WT for feature extraction of vibration signals and an application example in the field of gear boxes is given. The authors point out a lack of practical applications. As Sun et al. [195] state, an appropriate wavelet has to be selected to ensure optimal fault detection. Peng and Chu [21] give an overview of the development of wavelets and review the application of wavelet transform in condition monitoring and fault diagnosis. According to the authors, FFT is the most popular method. When dealing with non-stationary signals, the authors suggest WT for machine diagnostics. Recently, Yan et al. [196] review the application of wavelet transform in fault diagnosis for rotating machinery more. The authors identify the need to create new wavelet functions, since defect-related fault features are better extracted with wavelet functions similar to the signal.

The continuous wavelet transform solves the time-frequency resolution problem by sweeping the window size. A basis function ψ called mother wavelet (MW) is dilated and translated. This is the important difference to STFT, where the window width is fixed. The equation of CWT is

$$CWT_{\tau,s,\psi}(x) = \int_{-\infty}^{\infty} x(t)\,\psi^*_{\tau,s}(t)\,dt, \tag{4.3}$$

where $*$ stands for the conjugate and

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right). \tag{4.4}$$

The factor $1/\sqrt{s}$ normalizes the mother wavelet ψ. The transformed signal is dependent on the translation parameter τ and the scale parameter s. The parameter τ describes the translation of the window, accountable for the time position, the parameter s is proportional to the inverse of the frequency $s \sim 1/f$. It constrains the time resolution as well as the frequency resolution. The higher the number of s, the wider the window gets and global or lower frequencies are detected. A smaller value of s leads to a faster variation of the wavelet, so higher frequencies of the signal are detected. All windows are shrunk or widened versions of the MW. The shortcoming of CWT is a large increase in the number of data points. This leads to high computational load. A detailed interpretation of the parameters and the mathematical conditions for the existence of wavelets are given in [193].

Augner and Flandrin [197] evaluate the use of wavelet transform in multi-scale analysis of a signal through dilation and translations. The authors claim that WT was able to extract time-frequency features of a signal effectively and therefore the wavelet transform was more suitable for the analysis of non-stationary signals. Smith et al. [198] compare the performance of Haar, Morlet, and Daubechie wavelets on vibration detection. Albadour et al. [199] use the CWT on vibration signals and claim it would be an effective method in fault diagnosis, but point out the importance of selecting an appropriate mother wavelet.

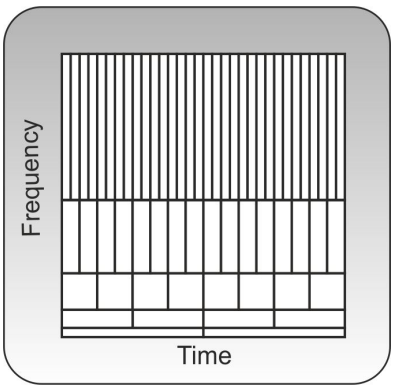


Figure 4.8: CWT resolution as a function of filter window width [19]

To execute the CWT, the MATLAB®toolbox for wavelets is used. The function for continuous wavelet transform *cwt* computes the coefficients of an input signal at real positive scales. The wavelet can be set individually, real or complex. Lee et al. [10] point out, that no systematic scheme has been developed to select a suitable wavelet. The selection has to be done based on expert judgment. In the present approach, the symmetric Morlet wavelet is used to avoid a weighing of certain components. An example of the application of a Morlet wavelet is shown in Figure 4.9. Low scales cause a compressed wavelet, illus-

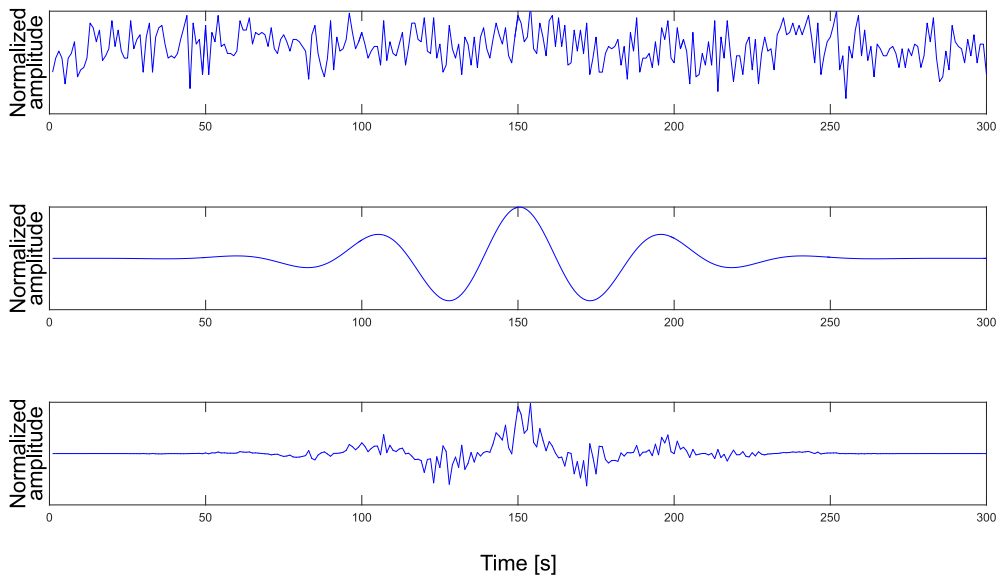


Figure 4.9: Application of a Morlet wavelet window function to a time signal; top: Raw time signal; Middle: Morlet window; bottom: Filtered Signal

trating the high frequency components. High scales set up a stretched wavelet, capturing low frequency components.

Discrete Wavelet Transform

In the early 1980s, Strömberg [200] dealt with the mathematical foundations for the use of discrete wavelets. Compared to CWT, the resulting number of data points can be reduced by calculating a low number of values corresponding to a lower time resolution so that a multi-scale analysis is possible. This is achieved by the discrete wavelet transform (DWT). The general mathematical expression is as follows

$$DWT_{m,k}(x) = \int_{-\infty}^{\infty} x(t)\,\psi_{m,k}(t)\,dt, \tag{4.5}$$

where

$$\psi_{m,k}(t) = \frac{1}{\sqrt{{s_0}^m}}\psi\left(\frac{t - k\tau_0}{{s_0}^m}\right). \tag{4.6}$$

The wavelet ψ is translated by $k\tau_0$ and scaled by ${s_0}^m$. In difference to CWT, the indices m and k are positive integers, so that the evaluation is completely discrete. Using DWT, the signal passes a filter belt (series of filters of different cutoff frequencies), which splits

the signal into its frequency components. In practice, the parameters $s_0 = 2$ and $\tau_0 = 1$ are chosen. The signal is split in the middle of the frequency band by high-pass and low-pass filters repeatedly until the width of the window is reached. The filtered signals contain redundant information. According to Nyquist's rule, downsampling by factor of 2 is allowed without aliasing [201]. This is visualized in Figure 4.10. The computational load can be reduced considerably by the use of downsampling [202].

Let the impulse response of the low-pass filter be g. The filter algorithm can then be written as

$$y\left[n\,t_0\right] = (x * g)\left[n\,t_0\right] = \sum_{k=-\infty}^{\infty} x\left[k\right] g\left[n\,t_0 - k\right], \tag{4.7}$$

where $(x * g)$ is the convolution operation and square brackets denote the row character of signals. The output from the low-pass filter has lost the detail information. The resulting values are called approximation coefficients. To preserve the detail coefficients, simultaneously a high-pass filter h is applied. The high pass and the low pass filter have to build a quadrature mirror filter, where $h\left(e^{j\Omega}\right) = g\left(e^{j(\pi-\Omega)}\right)$ holds. The parameter Ω is the normalized cutoff frequency, in this case $\Omega = \pi/2$. This condition is held by orthogonal wavelets, for instance the Daubechies wavelets [192].

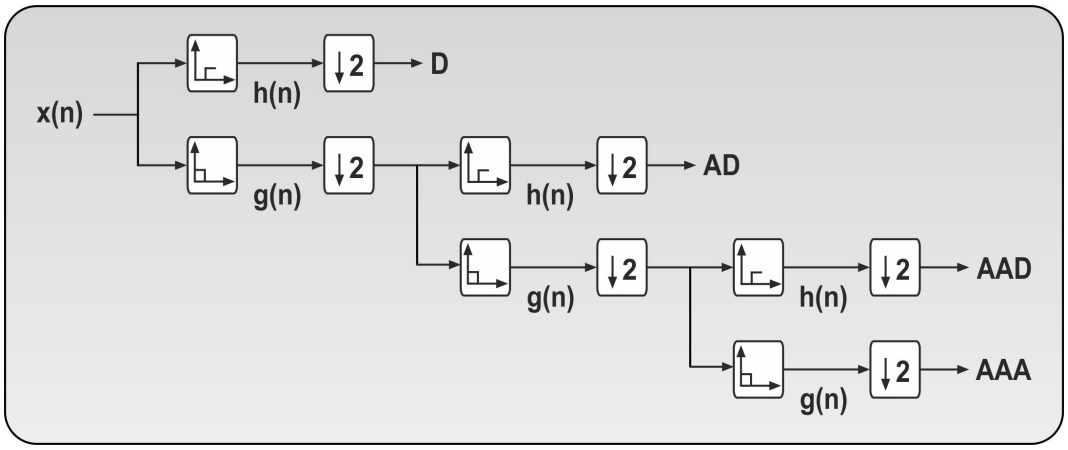


Figure 4.10: DWT wavelet filter belt [19]

In Figure 4.10, a wavelet filter belt is illustrated, where $x(n)$ denotes the input signal, $h(n)$ the high pass filter, $g(n)$ the low pass filter, D the detail component (high frequencies), and A the wavelet coefficient of the approximation component (low frequencies). This is repeated several times so that the details of the first approximation coefficient are noted as AD et sequentes.

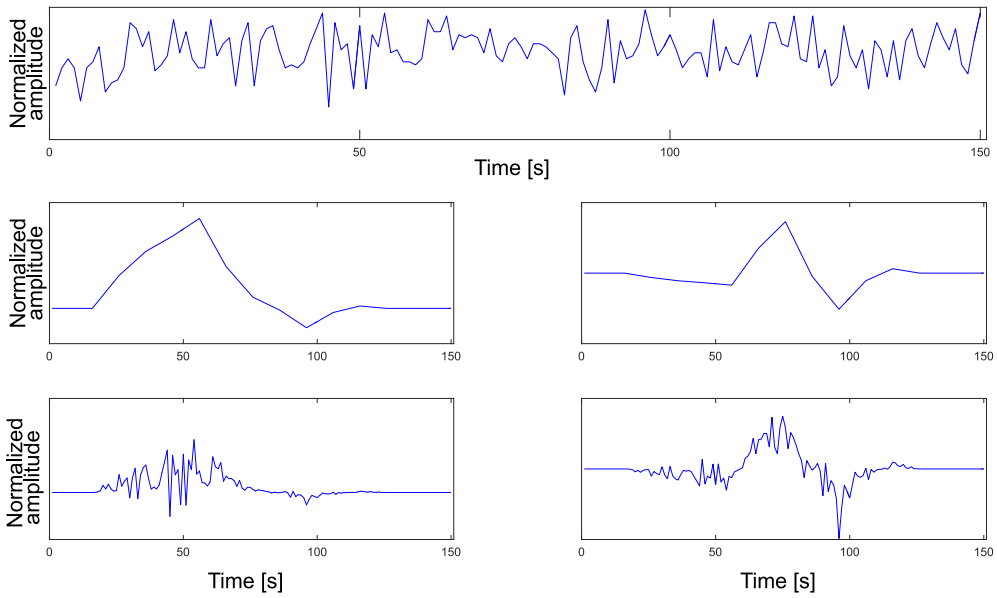


Figure 4.11: Wavelet window function; top: Raw signal; middle left: Scaling function; middle right: Mother wavelet; bottom: Filtered signal

Yao et al. [203] investigate online chatter detection and identification based on wavelet transform and a support vector machine during milling. For feature extraction from experimental data, DWT and wavelet packet transform (WPT) are used. The authors define three machine states as classes. For each class 10 training data sets and 5 test data sets are used. The authors claim the method would be robust for different machine conditions during the milling process with a detection rate of 95%. The work of Luczak et al. [204] compares the application of DWT, CWT, and DWT-FT to detect the resonance frequencies resulting from a mathematical simulation of a direct drive. Resonance frequencies could be detected using WT. The CWT is redundant and generates a high computational load. The DWT-FT allows to identify the mechanical resonance frequency components.

Variations of the standard DWT methods are used by a number of authors [205, 206, 207]. Cai et al. [205], e.g., suggest a sparsity-enabled decomposition method for feature extraction based on tunable Q-factor wavelet transform (TQWT), morphological component analysis (MCA), and split augmented Lagrangian shrinkage algorithm (SALSA) . By nonlinear decomposition, the proposed method exploits information on different oscillatory components. The merits of the new method have been verified by simulated and practical gearbox vibration signals. This application shows, that WT is suited to detect vibrational components of non-stationary signals in a noisy environment. Also, wavelet transform is commonly used for data compression and image processing [206, 207].

For DWT, the MATLAB®toolbox for wavelets is used. A Multilevel 1-D wavelet decomposition is performed on the input signals using *wavdec*. This function decomposes an input signal into its approximation and detail vectors. The highest decomposition level is computed for each particular wavelet. Also, wavelet decomposition filters can be set. In the present case, a Daubechie wavelet is used, since no further mathematical conditions have to be fulfilled using these orthogonal wavelets. The order of the wavelet does not give remarkable changes in the results. Therefore, the simplest - appart from the Haar wavelet- is used, a DB2 wavelet. The wavelet consitsts of two parts, a scaling function and a mother wavelet. The scaling function acts as the low pass filter and the mother wavelet as the highpass filter. The effect of the two functions applied as a window to a time signal is shown in Figure 4.11.

Wigner-Ville Distribution

In 1932, Wigner developed a distribution to apply quantum corrections to classical statistical mechanics. Ville identified the same function as quadratic representation of a signal's local time-frequency energy in 1948. Today, several names are used for this algorithm, a common one is Wigner-Ville Distribution, which is defined by

$$WVD_x(t,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x\left(t+\tau/2\right)\, x^*\left(t-\tau/2\right) e^{-j\omega\tau} d\tau. \tag{4.8}$$

The WVD is calculated for each point represented by the data triplet of signal x, time t, and frequency ω. In analogy to STFT, a window is shifted over the signal. In WVD the signal itself is simultaneously shifted into the opposite direction. The window x^* is the complex conjugate of the original signal. This comparison of the signal's information with its own information at another time has a structural resemblance to an autocorrelation function modified by the phase shift function $e^{-jw\tau}$.

This time-frequency energy distribution is a Cohen's class member, meaning it is covariant under translation in time and frequency. The interpretation of transformation results of real signals (non complex) is difficult so that the signal has to be expanded by its analytic associate given by the Hilbert Transform (HT) of the same signal as imaginary part. Basically, the HT gives the possibility to establish a relation between real and imaginary part of the Fourier transform of an analytical signal. Applied to general time functions, the transform is given by the following equation

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau, \tag{4.9}$$

where $x(\tau)$ is the original time-dependent signal and $y(t)$ denotes the Hilbert-transformed

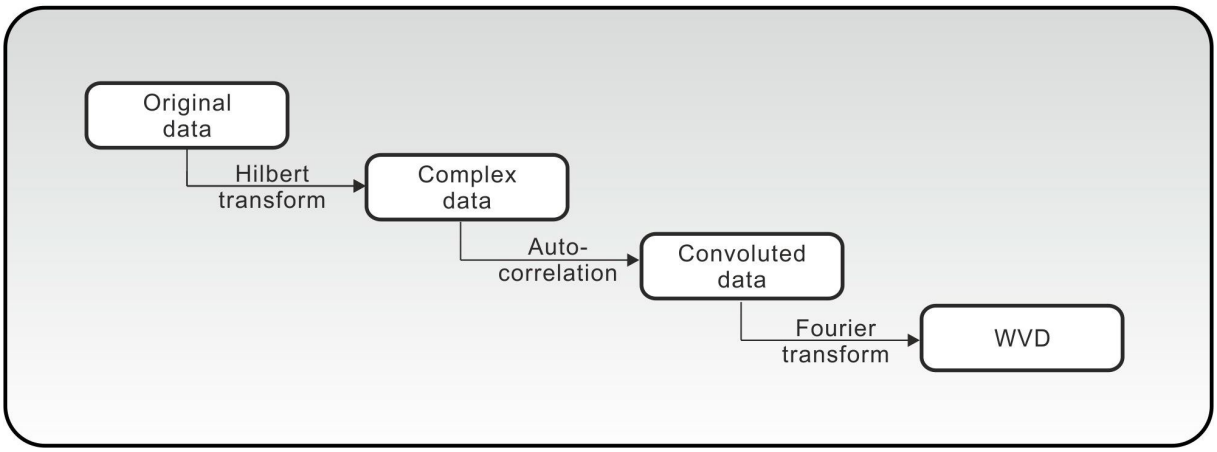


Figure 4.12: Visualization of a Wigner Ville distribution

signal. To avoid the singularity at $t = \tau$, the Cauchy principal value of the integral indicated by P is evaluated, this allows to calculate the value of the integral. An overview of the steps of a WVD is given in Figure 4.12.

A specific property of the transform are negative output values, which do not correspond to physically meaningful results. In the original application of the transform to quantum mechanics this property is no hinderance. As a quadratic function, it will lead to interference terms, which will mislead the analysis and have to be considered carefully [21]. The reduction of the interference terms is achieved by averaging. This means low-pass filtering and leads to a loss of time-frequency resolution. This has to be kept in mind during interpretation.

These shortcomings of the method are substantial when applied to multi-frequency signals. Therefore WVD is rarely used in applications concerning such data.

Application examples to experimental data can be found at Lamraoui et al. [208] and Climente-Alarcona et al. [209]. Lamraoui et al. use WVD for a cyclostationary approach for monitoring chatter and tool wear in milling. It is applied to experimental accelerometer data. According to the authors, the results show that Wigner-Ville spectrums are useful parameters for early diagnosis. Climente-Alarcona apply WVD for the detection of rotor asymmetries and eccentricity through high-order harmonics. In both cases, the sought features are successfully detected. The Hilbert transform of the original data does not show any characteristic pattern. Therefore the graph of the Hilbert transform is not show here.

Before calculating the WVD, a Hilbert transform is performed using *hilbert*. The real data are transposed to an analytic signal. Additionally, a plausibility check is performed. For the MATLAB® application a column vector is needed. This ensures that the trans-

lation parameter tau is shifted in the right direction. The signal component at $t + \tau$ is multiplied with the complex conjugate of the signal component at $t - \tau$. After that, a standard MATLAB® FFT is executed using fft to calculate the WVD. To cancel negative interference terms in the result, absolute values are regarded.

Empirical Mode Decomposition

Basically, the HT establishes a relation between real and imaginary part of the Fourier transform of an analytic signal. In 1998, Huang et al. [31] developed the idea to extend the application of HT to non-analytical signals. Hereto, the signals are decomposed into components which are sufficiently analytic. The decomposition is detailed in the following passage.

Non-stationary signals contain multi-frequency components, so that the relation between frequency (retrieved from real part of FT) and phase (retrieved from imaginary part of FT) does not hold. The Empirical Mode Decomposition decomposes any given signal into Intrinsic Mode Functions (IMF). The IMF are treatable as mono-component functions, so that the HT relation between phase $\varphi(t)$ and frequency $\omega(t)$ holds approximately. A recently published review on EMD is given by Lei et al. [210]. Here, theoretical drawbacks of EMD are evaluated, stating that no practical disadvantages are observed examining fault diagnosis of rotating machinery.

An IMF represents specific kinds of oscillation modes of the original signal. Following the definition of Huang, the IMF has to satisfy two conditions "[...] the number of extrema and the number of zero crossings must either equal or differ at most by one" in the whole data set and "the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero" [31]. These constraints are visualized in Figure 4.13. The number of IMF necessary to reconstruct the original signal is finite and often small. Bisu et al. [211] note an experimental approach on dynamic analysis for monitoring and diagnosis of a milling process. The authors apply EMD followed by HT, which is called Hilbert-Huang transform (HHT), using the described envelope method to identify the dynamic behavior of a milling process. The values of warning and alarm thresholds are determined considering the optimal machine performance.

Georgoulas et al. [212] combine EMD and SVM for anomaly detection in rotating machinery. For feature extraction, EMD is used, and selected IMF are transferred to three different anomaly detectors. The test data set provides four different load conditions. Eleven data sets are used. The results show that all fault states can be detected without false alarms using an attribute bagging scheme and all three detectors.

Peng et al. [213] give the following algorithm for EMD.

EMD Algorithm [213]*:*

(1) Initialize: $r_1 = x(t)$, and $i = 1$

(2) Extract the i$^{\text{th}}$ IMF

(a) Initialize: $h_{i(k-1)} = r_i, k = 1$

(b) Extract the local extrema and the minima of $h_{i(k-1)}$

(c) Cubic spline interpolation of local extrema from upper and lower envelopes of $h_{i(k-1)}$

(d) Calculate the mean $m_{i(k-1)}$ of the upper and lower envelopes of $h_{i(k-1)}$

(e) Let $h_{ik} = h_{i(k-1)} - m_{i(k-1)}$

(f) If h_{ik} is an IMF then set $IMFi = h_{ik}$, else go to step (b) with $k = k + 1$

(3) Define $r_{i+1} = r_i - IMFi$

(4) If r_{i+1} still has least 2 extrema then go to step (2) else decomposition process is finished and r_{i+1} is the residue of the signal

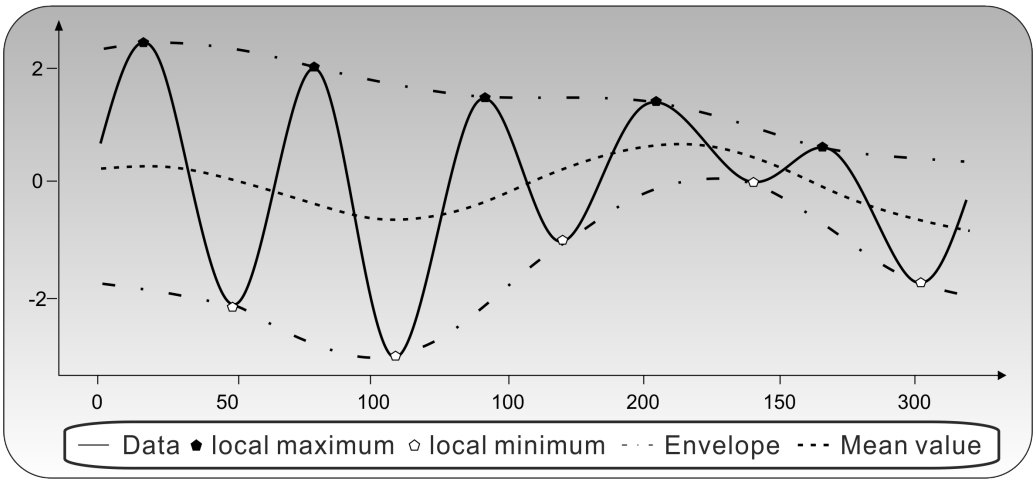


Figure 4.13: EMD constraints [19]

The essential MATLAB®functions needed to perform the EMD are cubic spline interpolation *spline* and peak search *findpeaks*. The maxima and minima in the input data are located by the appropriate MATLAB®function. The maxima respectively the minima are interpolated by the MATLAB® cubic spline function. The two mathematical conditions for EMD are that the mean value has to be zero and the amount of maxima and minima must not differ by more than one. These are tested in two iterative loops, resulting in a vector holding the intrinsic mode function.

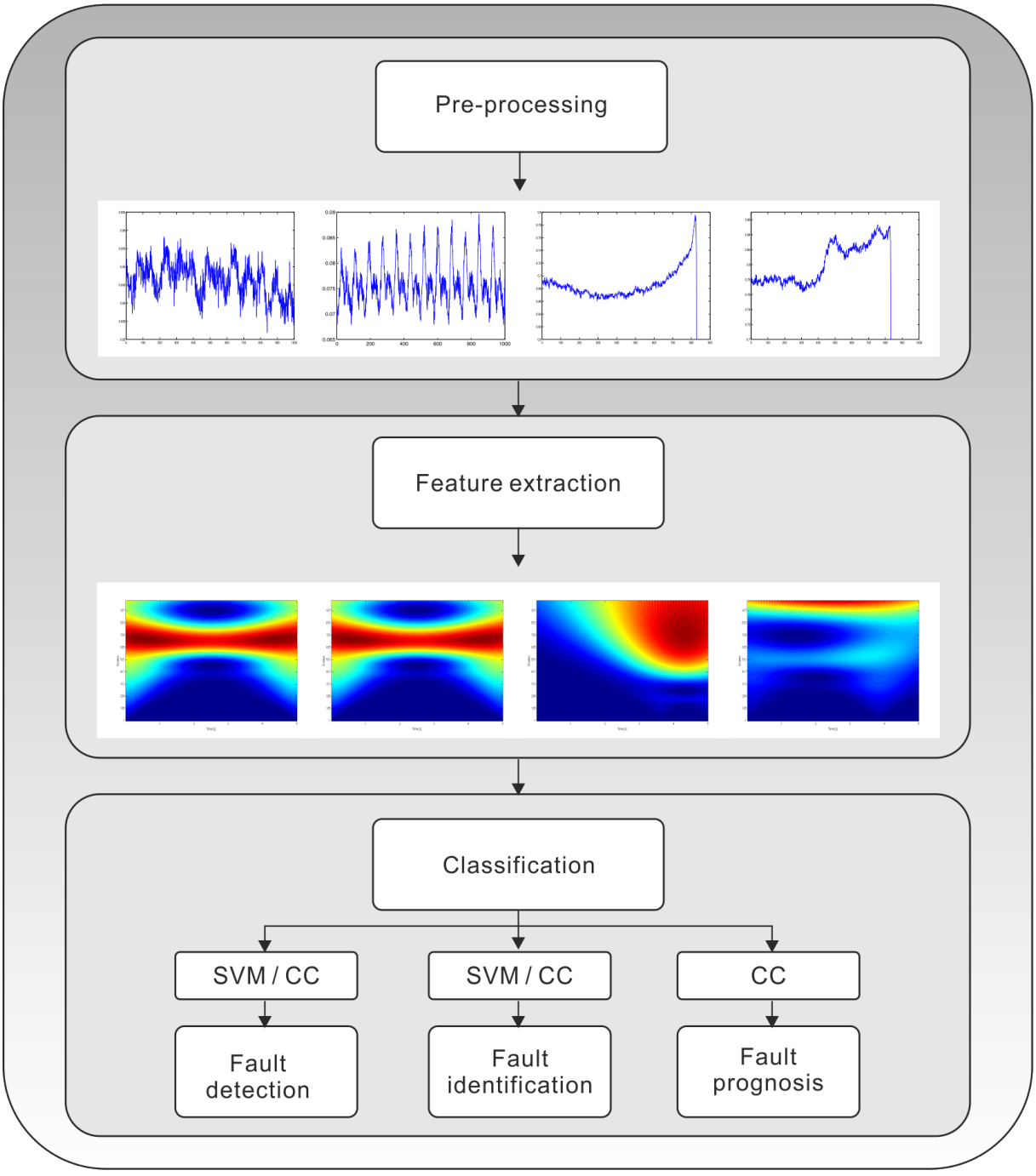


Figure 4.14: Visualization of signal processing steps

4.4.3 Processing techniques for classification

The results of the feature extraction step have to be arranged in groups coresponding to the machine states. In the basic concept, a suited classifier has to be picked out. In this approach, a self-learning algorithm SVM is used. Additionally, in the case of EMD, the correlation coefficients of a cross correlation are used for thresholding to classify the system state. The combination of these methods has not been found in the literature and is a *new signal processing* technique presented in the present work. In this section, a basic theoretical background is given and the set-up for three different classification tasks are discussed: 1) the fault detection, 2) the fault identification, and 3) the fault prognosis. The combination of the signal processing steps is visualized in Figure 4.14. The focus lies on the pre-processed time signal and the effect of the feature extraction, in this case exemplarily performed by CWT. This method has been choosen for its graphical representation that shows an eye-catching impression of the effect of feature extraction. The results of the classification are presented in detail in Chapter 5.

Support Vector Machine

The Support Vector Machine algorithm realizes the classification of data. It is trained with prepared data sets of known classes (training data) to distinguish between certain patterns. The trained model is used to classify unknown data (test data). Abe [214] gives a comprehensive resume. The training data are given in vector-form

$$(y_1, x_1) \dots (y_n, x_n)\, x \in \mathbb{R}^n, y \in \{-1, +1\}\,. \tag{4.10}$$

The linear SVM can be used on easily spreadable classes. The hyper plane is defined as

$$\mathcal{H}\,(w, b) := \left\{\forall x | w^T x + b = 0\right\}, \tag{4.11}$$

where w is the normal vector to the hyperplane and b is the parameter needed to calculate the normal distance of the hyperplane to the origin. The Parameters in Equation 4.11 are scalable, meaning that

$$\mathcal{H}\,(w, b) = \left\{\forall x | cw^T x + cb = 0\right\} \tag{4.12}$$

with $c \in \mathbb{R}$ and $c \neq 0$ leads to the same hyperplane as Equation 4.11.

A unique hyperplane can be defined by scaling the parameters w and b to fulfil the condition

$$\min_{x_i} |w^T x_i + b| \stackrel{!}{=} 1 \tag{4.13}$$

for the vectors x_i of the training data set. Such a hyperplane is called canonic hyperplane.

The euclidean distance of a point x_i to the hyper plane is

$$d(\mathcal{H}; x) := \frac{|w^T x_i + b|}{||w||}. \tag{4.14}$$

The data poits x_i closest to the hyperplane are the support vectors. The distance of the support vectors to the hyperplane shall reach a maximum. This distance, the margin ζ is calculated by

$$\zeta(\mathcal{H}) = \min_{x_i} d(\mathcal{H}; x_i) = \frac{1}{||w||} \left[\min_{x_i} |w^T x_i + b| \right] = \frac{1}{||w||}. \tag{4.15}$$

To maximize ζ, $||w||$ hast to be minimized. The following optimization problem has to be solved

$$\Theta(w) = \arg\min_{w,b} \left[\frac{1}{2} ||w||^2 \right]. \tag{4.16}$$

To achieve this without violating the condition of a canonic hyper plane (Eq. 4.13), the constraint

$$y_i \left(\left(w^T x_i \right) + b \right) \geq 1, \; i = 1, \ldots n \tag{4.17}$$

has to be fulfilled. This optimization problem with constraints is solved using the method of Lagrange multipliers, resulting in

$$\bar{w} = \sum_{i=1}^{n} \bar{\alpha}_i y_i x_i \tag{4.18}$$

for $\bar{w}$, with $\bar{\alpha}_i$ being the Lagrange multipliers, and

$$\bar{b} = -\frac{1}{2} \bar{w} \left[x_r + x_s \right] \tag{4.19}$$

for $\bar{b}$, where the indices r and s indicate the support vectors and

$$\bar{\alpha}_r, \bar{\alpha}_s > 0, \; y_r = 1, \; y_s = -1. \tag{4.20}$$

The optimal separating canonical hyper plane results in

$$f(x) = \text{sign}\left(\bar{w}^T x + \bar{b}\right). \tag{4.21}$$

In practice, non-spreadable single data points will appear in the wrong class and may lead to malfunction of the algorithm. A certain number of wrongly classified data points is allowed within a soft margin. Data points on the wrong side of the hyper plane are measured with ξ_i. Therefore, Θ is changed to

$$\bar{\Theta} = \Theta + C\sum_{i=1}^{n}\xi_i, \tag{4.22}$$

where

$$y_i\left(x_i^T w + b\right) \geq 1 - \xi_i,\ \xi_i \geq 0. \tag{4.23}$$

Figure 4.15 illustrates the optimization problem exemplary. Two features are to be separated by an optimal hyperplane, giving a maximum margin to the nearest data point.

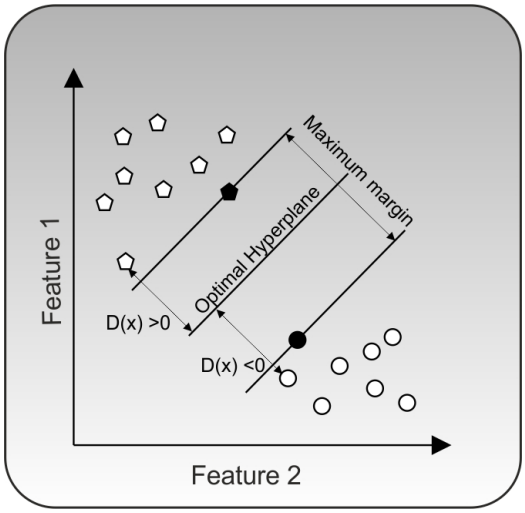


Figure 4.15: SVM hyperplane separation with maximum margin

In the present approach, a non-linear SVM is used with a kernel function that may transform the data from the inputspace I to a suited feature space $\mathcal{F}$ in which a separation of the classes is possible. The SVM is applied to the features extracted with STFT, CWT, DWT, WVD, and EMD. For the realization of the SVM, the open source toolbox LIBSVM [215] programmed by Chih-Chung Chang and Chih-Jen Lin is used.

A grid search varying the values of two parameters defining the used kernel is performed to achieve best classification accuracy. Over-fitting of the kernel may lead to bad performance of the classifier, because the restrictions for the identification of the classes becomes too narrow. To avoid over-fitting, the second best results of the grid search are used. Due to the feature extraction step, some of the data are arranged in matrix form. This applies to the output of STFT, CWT, and WVD. The output of DWT and EMD is arranged in several linear vectors. For the application of SVM, all output data sets are rearranged to form a single vector. The training and test data sets contain each a label vector with true classes. The training and test data are stratified. When applying cross correlation, it is ensured that no signal is in both groups in the same test cycle.

Cross-correlation

The cross-correlation measures the similarity of two given inputs signals. The correlation function is

$$R_{xy}(\tau) = \lim_{T_F \to \infty} \frac{1}{T_F} \int_{-T_F/2}^{T_F/2} x\,(t)\; y\,(t+\tau)\, dt, \tag{4.24}$$

where R_{xy} is the correlation function depending on the time lag τ, where t represents the time, and T_F the time window. The input signals $x(t)$ and $y(t)$ are swept along the time axis and thus compared to each other. An accordance between the signals can be assumed for positive values of R_{xy}. Higher values indicate stronger, smaller values weaker similarity, and values of R_{xy} near zero show the absence of a connection between the input signals. Vice versa, for negative values of R_{xy} an opposing connection can be assumed. An example to illustrate the effect of CC is shown in Figure 4.16, where two rectangular signals with the time difference 100 are used. The result of CC shows the maximum, in this case $R_{xy} = 1$, at time value 100, corresponding to the time difference of the signals.

Applied to the features extracted by EMD, the results of a CC are used to distinct the level of symmetry of the IMFs of a signal. Hereto, the input signal is split into half. The second signal half is mirrored, and the two half-signals are cross-correlated. The amplitude and the position of the maximum of the correlation coefficients gives information about the symmetry of input signals. High symmetry will lead to a value near one at time position zero. One or more IMFs may show this behavior. Cross correlation is applied to all of them and the highest correlation factor is used to discriminate the classes. The thresholds for the different classes can be varied. A high value may miss a regular condition leading to high rate of false alarms. A low value may rate a faulty conditions as regular condition leading to false negative results.

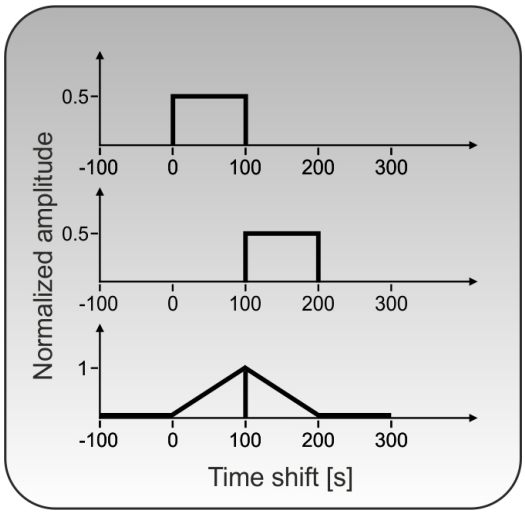


Figure 4.16: Cross-Correlation of two input signals

4.4.4 Classification for fault detection

At the vary beginning of an event, the reaction of machine operators on upcoming cobbles and shearing tails is similar. It would already be useful to know about the deviation during the production process. Therefore, the first classification task of fault detection is meant to distinguish two general system states: a regular state without deviations and a deviated system state. This is of great importance for the practical application to preserve the machine aggregates.

In matters of fault detection, all sample data sets are used. The initial four classes are summarized as two new classes. The regular State 1 with strip in stand without known deviation and the regular State 2 without strip in stand without known deviation are summarized as one class. The deviated State 3 and the deviated State 4 are summarized as the second class. A new label vector is generated mirroring the new classes. The feature extraction with STFT, CWT, DWT, WVD, and EMD is executed for each data set individually. The resulting features are used as input for the classifiers. Using the data base broadened by cross-validation, 320 data sets with 160 data sets per class are available. For training, 240 data sets with 120 data sets per class are used. The remaining 80 data sets with 40 per class are used for testing. The feature extraction results of the five methods are used as input for the support vector machine. The feature extraction results of EMD show a particularity that lead to the idea of symmetry considerations for fault detection. Therefore, the feature extraction results of EMD are used as input for the cross-correlation. The correlation coefficient is thresholded to derive the system state.

4.4.5 Classification for state identification

In general, it is desirable to distinguish all of the machine states. This means that the states are no longer combined to two classes but treated separately. The four classes are regular without known deviation with strip in stand, regular without known deviation without strip in stand, deviated by cobble, and deviated by shearing tail. The task is to distinguish the four states, especially the two deviated states.

For the task of state identification, all sample data sets are used. The label vector is not changed, it distinguishes betwen four classes. The feature extraction with STFT, CWT, DWT, WVD, and EMD is executed for all data sets. The resulting features are used as input for the classifier. Here, the data base broadened by cross-validation contains 320 data sets with 80 data sets per class. For training, 240 data sets with 60 data sets per class are used. The remaining 80 data sets with 20 per class are used for testing. The extracted features are used as input for a SVM. Additionally, the newly developed classification method is applied. The symmetry degree of the IMF is regraded to classify the system states by cross-correlation coefficients.

4.4.6 Classification for fault prognosis

Fault prognosis describes the prediction of a deviated system state. Based on the evaluation of the classification rates, the newly developed approach combining the EMD with CC is used, because it leads to the best results regarding fault detection rate as well as a low rate of false alarms.

For the prognosis task, a detection rate of fault detection and fault identification are evaluated. For the detection rate, the two states without known deviation are summarized as one class and the two states with deviation are summarized as the second class. This leads to 240 training data sets with 120 per class and 80 test data sets with 40 per class. The detection of an upcoming fault is of vital importance for production process, because it can avoid severe damages. Additionally, the identification of the fault avoided by the detection will be helpful for quality management purposes. To obtain the detection rate of fault identification, all four states are differentiated. This leads to 240 training data sets with 60 sets per class and 80 test data sets with 20 per class. To predict of the state development, the original set length is cut from five to three seconds in advance of the fault. The newly introduced method combination of EMD with CC is applied to the shortened vector. The amplitude of the cross-correlation coefficients is thresholded. This threshold gives information on the symmetry degree of the input signal and enables a classification.

4.5 Validation

The results are validated by statistical analysis. Several statistical methods can be applied depending on the kind of analysis. Notably, there are two kinds of statistical analysis: the dependence analysis and the independence analysis. In this approach, the dependence of not normally distributed variables is to be examined. The χ^2-test is the method of choice for this task. The application of the χ^2-test is legitimate, since the number of data sets is high enough [216]. Two additional thumb rules are given as prerequisites for the test [216]. Depending on the sources, the minimal values for the expected class counts are demanded to be higher than values between 1 and 5. In this application, this condition is not met in the case of STFT-SVM. A correction after Yates could be applied, leading to a more conservative interpretation. An over-correction may fail to reject the null hypothesis. Therefore, this correction has not been applied here. Additionally, a McNemar's test is performed to determine whether paired samples are interrelated or not. When other statistical tests require independence of the tested observations and cannot be applied to correlated data, McNemar's test has been developed especially for this task. A correction according to Edwards has been applied to the McNemar's test to yield conservative results.

4.5.1 χ^2 test

The χ^2 hypothesis test compares a statistical model of the data to the observed data. Two statistical models are chosen here to be compared. The first null hypothesis $H_{0.1}$ is assuming a normal distribution or random occurrence for the results. This assumption means that the mathematical method applied to the data is not able to identify the machine states at all but classifies randomly. The alternative hypothesis $H_{1.1}$ is stating that the respective method is performing not randomly. If the null hypothesis can be significantly rejected, the alternative hypothesis $H_{1.1}$ is chosen, meaning that the tested distribution is not random. If the null hypothesis cannot be rejected, this does not automatically lead to the acceptance of the null hypothesis. The only possible statement is that it cannot be proven that the method is performing otherwise than randomly classified results. In Nachtigall et al. [217], the test formula is given as

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_{o,i} - f_{e,i})^2}{f_{e,i}}. \tag{4.25}$$

In this context, f_o are the observed frequencies and f_e the expected frequencies. The parameter k gives the number of possible outcomes.

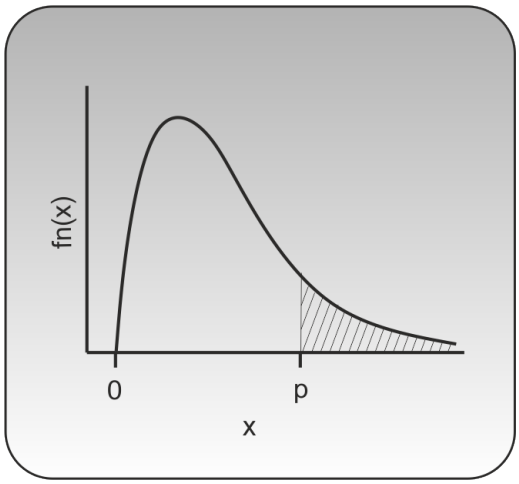


Figure 4.17: Density function of a χ^2 distribution with significance level p

The second null hypothesis $H_{0.2}$ is assuming that the results of the respective method will hit the correct class with a probability of 80%. The third null hypothesis $H_{0.3}$ is assuming a probability of 90%. The values of 80% and 90%, respectively, are chosen corresponding to the observed results. If the compared method performs better or worse, the respective null hypothesis will be rejected. The absolute values state if the method performs better or worse than 80% or 90% probability. A significance level of $p = 0,05$ is accepted as appropriate to reject the null hypothesis. Figure 4.17 visualizes the density of the χ^2 distribution for an exemplary degree of freedom. The cross-hatched sector shows the quantile of the significance level p.

4.5.2 McNemar's test

To validate the experimental results and to point out the non-random classification, a McNemar's test is performed [218]. A McNemar's test is a statistical method to determine whether paired samples are interrelated or not. The null hypothesis is that no variation takes place, consequently the alternative hypothesis is that a variation takes place. The original application of McNemar's test was to medical research. The expected outcome was to decide whether or not a difference after a medical treatment was statistically relevant. The test is applied here to the output of the different algorithms to decide if the results are significantly different. Examined are the numbers of data sets that are faultily or correctly classified by both compared methods. Two possible assumptions can be made on the results of the algorithms. The first is that all data sets faultily classified by one method are also faulty classified by the second method. The other assumption would be that both classifications are distinct. The first assumption leads to conservative numbers, assuming a greater similarity of the compared algorithms. Therefore, this option is chosen. As stated above, a continuity correction after Edwards [219] is executed to yield

a conservative result. The test is applied to the broadened data base. Thus, classes with less than five entries are avoided.

The formula of McNemar's test resembles to a χ^2-distribution with one degree of freedom [218]. The formula corrected after Edwards is

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}}, \tag{4.26}$$

where

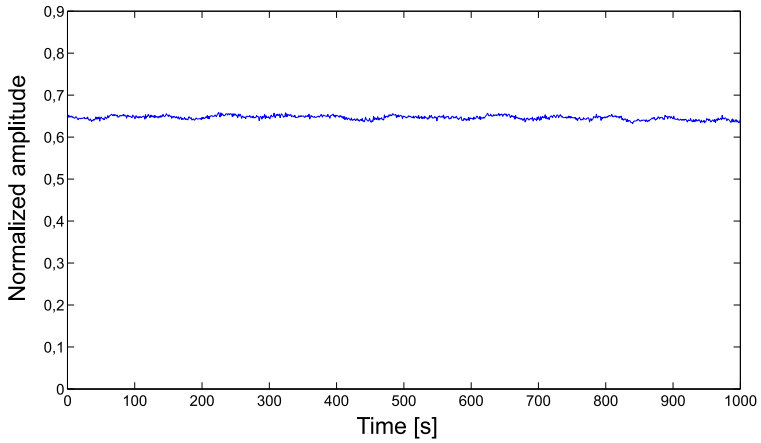
n_{00} := number of objects wrong classified by A and B
n_{01} := number of wrong classifications by A, which are true by B
n_{10} := number of wrong classifications by B, which are true by A
n_{11} := number of objects true classified by A and B.

5 Experimental results and validation

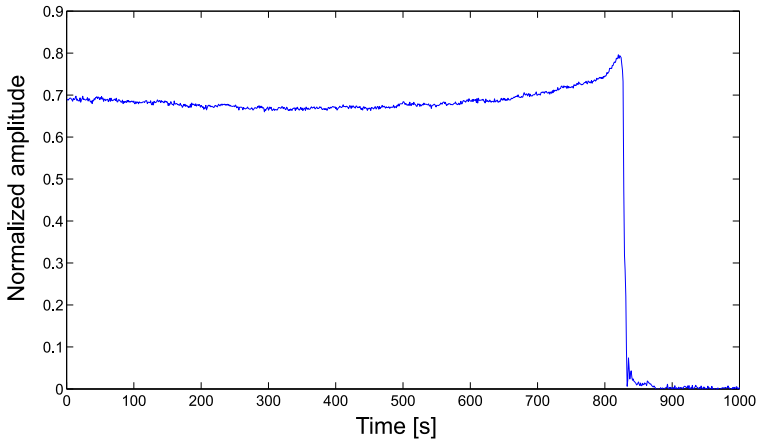
In this chapter, fault detection, fault identification, and prognosis will be applied to real data from a seven stand finishing mill to detect deviations in strip travel. In this context, the meaning of these terms is defined as follows. Fault detection means that a deviated system condition can be derived from the signal. Fault detection is important in applications where the deviation is not visible during the running process. If the fault can be detected by an automated system, possibly severe damages may be avoided. Fault identification will give further information on the type of deviation. The fault identification has to meet higher requirements. Therefore, the detection rate of fault identification is expected to be poorer than the detection rate of fault detection. The prevention of deviations and damages on the machine is desired. Fault prognosis will enable in-time reactions to prevent deviations. The content of this chapter is based on contributions already published [19, 179, 180, 181].

5.1 Fault detection

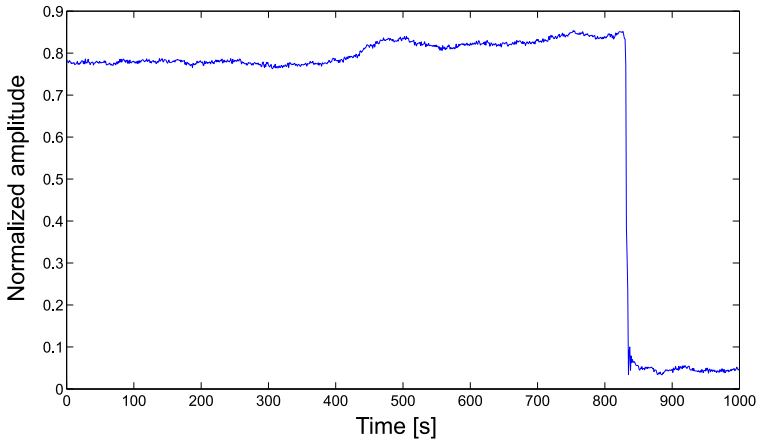
In this chapter, the mathematical methods previously introduced are applied to real data. This is followed by a detailed performance assessment. The fault detection rate is examined, meaning the probability to distinguish a regular from a deviating system state. The results are statistically validated. The data sets with a length of five seconds are visualized in Figure 5.1. The graph indicated by “a” is a fault-free state, meaning that the process with slab i stand is running without failure. The graph in the middle indicated by “b” is a fault case “cobble”, and the graph in the bottom indicated by “c” is a fault case “shearing tail”. The time signal of fault-free case does not show characteristic frequencies as has been tested in previous work [185]. The deviations from mean value are primarily randomly distributed, showing some vibrations that are not typical for the process and vary with different rolling conditions. The time signal of cobble (Fig. 5.1.b) shows an increase before the occurrence of the fault. From Figure 4.4, it can be argued that similar variations of the roll force difference also appear in fault-free states. Therefore the increase on its own is not an indicator of a cobble. The same argument is applicable to shearing tales (Fig. 5.1.c).



(a) Time signal of fault-free case



(b) Time signal of cobble



(c) Time signal of shearing tail

Figure 5.1: Graphical representation of pre-processed time signals

5.1.1 Graphical results of feature extraction

For feature extraction, the five selected time-frequency analysis methods (cf. Chapter 4) are applied. The graphical representation of the method's results for all four system states is given in Appendix A. Figures 5.2-5.6 show exemplarily the graphical results of three states. In all figures, the top figure shows the result for the fault-free case, system State 1, denoted by "a", the figure placed in the middle, denoted by "b", shows the result for the fault case "cobble", system State 3, and the figure at the bottom, denoted by "c" shows the result of feature extraction for the fault case "shearing tail", system State 4. The abscissa shows the index number of the result data. In case of time scales, as in STFT and CWT, this corresponds to 5 seconds.

Short-Time Fourier Transform

In Fig. 5.2, the results of STFT are illustrated in a time-frequency plane. The fault-free case is shown in Figure 5.2a, the fault case "cobble" is shown in Figure 5.2b , and the fault case "shearing tail" is shown in Figure 5.2c. The scaling of the amplitude (in arbitrary units) is the same in all three figures. The three system conditions show similar behavior up to the time of occurrence of the fault at about 3.2 seconds in case of cobble and 3.6 seconds in case of shearing tail. At these points in time, the STFT results show a strong deviation in both fault cases compared to the results of the fault-free case. A broadband distribution of frequency components is visible. The intensity decays to higher frequencies. This is a well known behavior of FT at sharp signal edges.

Continuous Wavelet Transform

The results of the application of CWT to the data sets are illustrated in Figure 5.3. The results are presented on a time-scale plane as scalogram. Low frequencies correspond to the upper edge of the graphical representations whereas higher frequencies correspond to the lower edge. The graphical representation of both fault cases (Fig. 5.3b and Fig. 5.3c) are clearly distinguishable from fault-free case (Fig. 5.3a) by machine operators or experts. For scale values of about 700, the fault-free case shows nearly continuous behavior, whereas the fault case "cobble" shows a strong increase in amplitude, and the fault case shearing tail shows a decrease at low times and an increase at higher time values. In both fault cases, starting at about 1.5 seconds, this change is visible before the occurrence of the fault. In contrast to STFT, the time of the fault occurrence known from the event-based data is not clearly visible.

Discrete Wavelet Transform

The graphical representation of the results using DWT (Fig. 5.4) shows the decomposition levels from higher to lower frequencies respectively from top to bottom. The abscissa scales give sample numbers, in case of the upper frame of the graphic from 500 to 1000. This corresponds to the number of data points in the high frequency part of the result, which is 500 due to downsampling. In each following subplot, the number of data points is reduced by a factor of two. In all subplots, the abscissa corresponds to the total elapsed time, regardless of the number of data points. In the fault-free case (Fig. 5.4a), essentially noise-like features can be seen in the first two decomposition levels. A high frequency glitch in the first decomposition level of the fault case "cobble" (Fig. 5.4b) indicates the occurrence time of the fault. The same applies to the fault case "shearing tail" (Fig. 5.4c). Compared to the fault-free case (Fig. 5.4a), the absolute values of the first seven decomposition levels of the two fault cases are at least one, mostly two orders higher in magnitude.

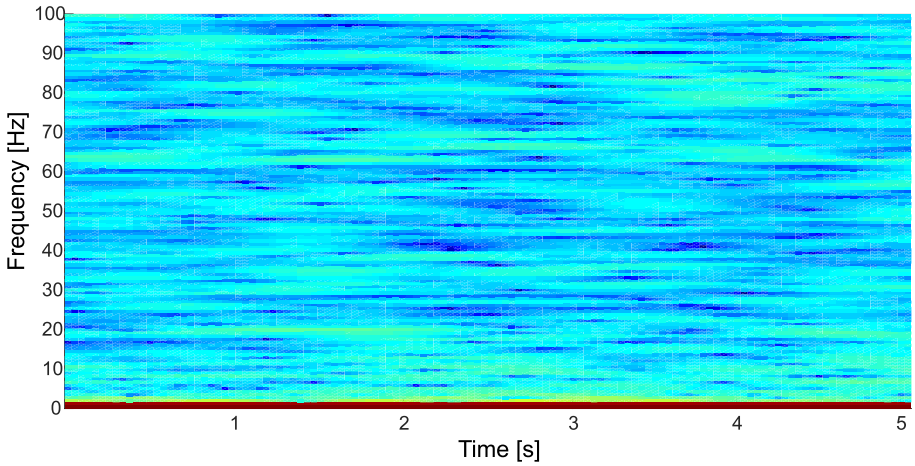
Wigner-Ville Distribution

The 3D plot in Fig. 5.5 shows the graphical representation of the results of the application of WVD to the data sets. Near 0 Hz and 100 Hz, contributions with values several magnitudes higher than in the middle part of the time-frequency plane appear. Therefore, the first and the last 100 data points along the frequency axis and for similar reasons the first and the last 20 data points along the time axis are omitted to show the behavior of the result in the middle part of the time-frequency plane. In the illustration, some features result from interference terms ot the non-linear algorithm. The results of the application to the fault-free case are represented in Figure 5.5a. The absolute values of the magnitude (arbitrary values) in the shown time-frequency area are less than five. The data of the fault case "cobble" in Figure 5.5b show higher amplitudes in the region of 2-3 seconds and around 4.5 seconds. The fast changes in amplitude indicate that these structures are results of interference terms. Obviously, they cannot be assigned to the occurrence of the fault. In the fault case, "shearing tails" in Figure 5.5.c, the data show additional strong oscillations for low frequencies. Again, these structures will have to be assigned to interference terms.

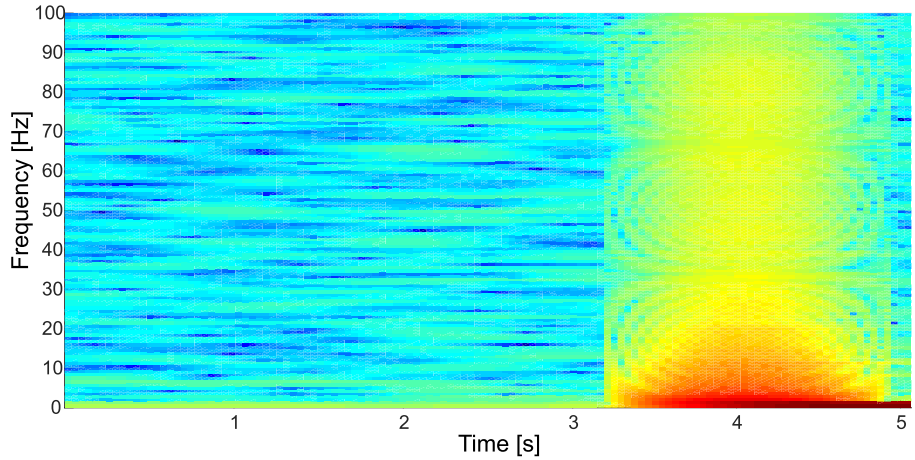
Empirical Mode Decomposition

At first glance, the graphical results of EMD as given in Fig. 5.6 seem to be similar to those applying DWT to the same data. The intrinsic mode functions shown span the whole time scale. There is no explicit frequency filtering in the algorithm. Instead, the IMF are empirically adapted oscillating modes of the time trend. The IMF in the first line of the graphic shows the detailed changes in the original time signal. Each following

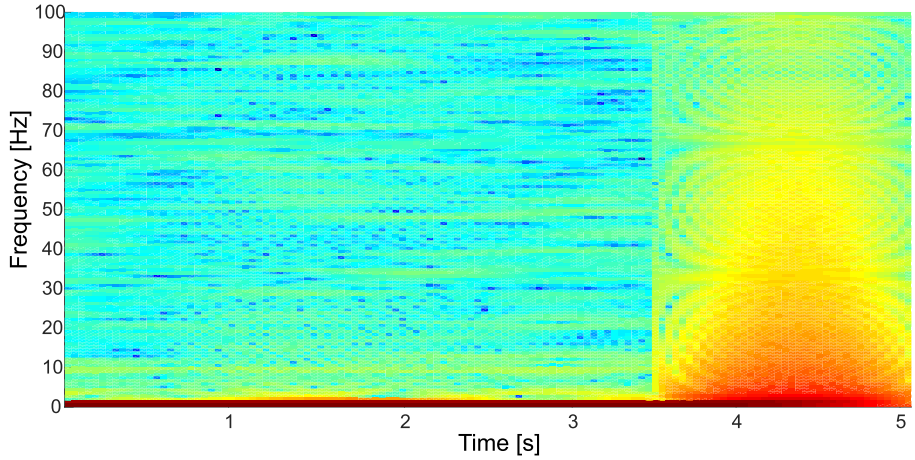
line gives the remaining time-dependent behavior of the signal until the last line shows the residual. The IMF's shapes of the fault-free case (Fig. 5.6a) are rather symmetrical. As with DWT, a glitch indicates the time position of the fault in fault case "cobble" in Figure 5.6b in the first, second and third IMF. In the fault case, "shearing tail", such glitches are visible in the first and second line of the graphic in Figure 5.6c. For both fault cases, the lack of symmetry is obvious.



(a) Application of STFT to fault-free case [19]

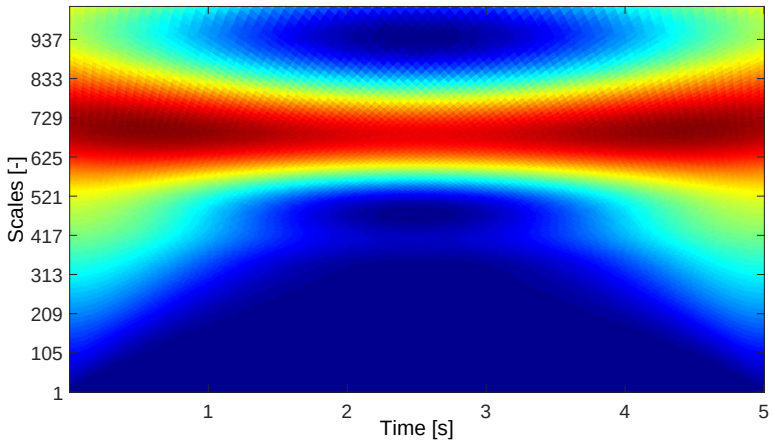


(b) Application of STFT to fault case (cobble) [19]

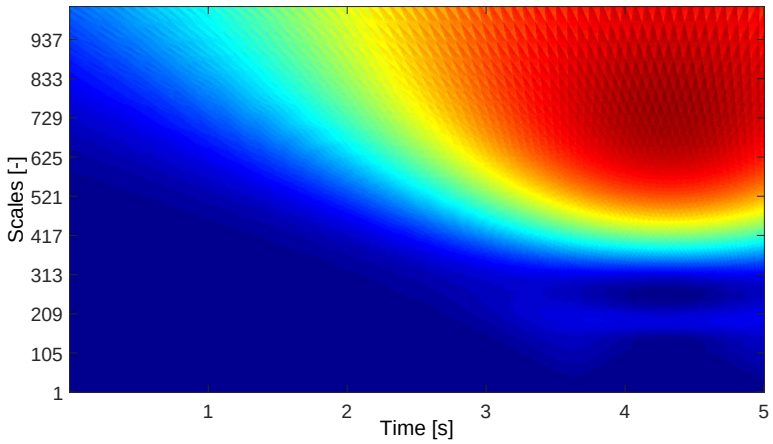


(c) Application of STFT to fault case (shearing tail)

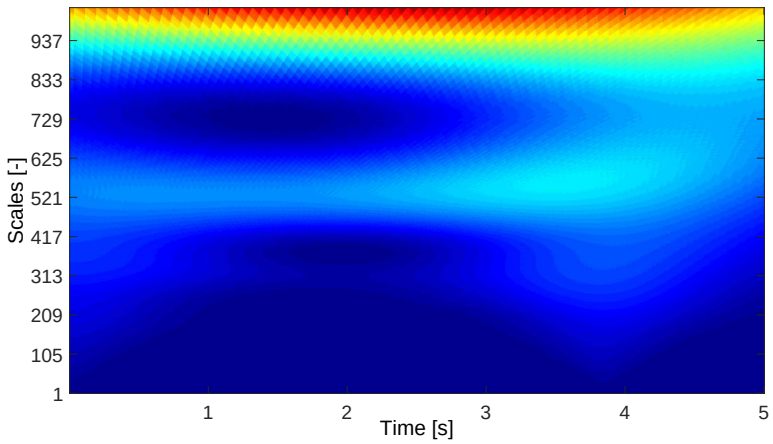
Figure 5.2: Graphical results of STFT



(a) Application of CWT to fault-free case [19]

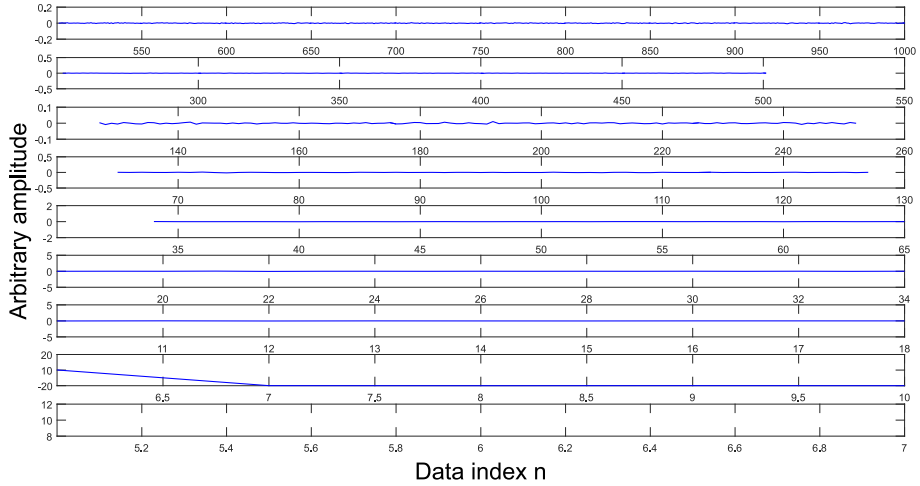


(b) Application of CWT to fault case (cobble) [19]

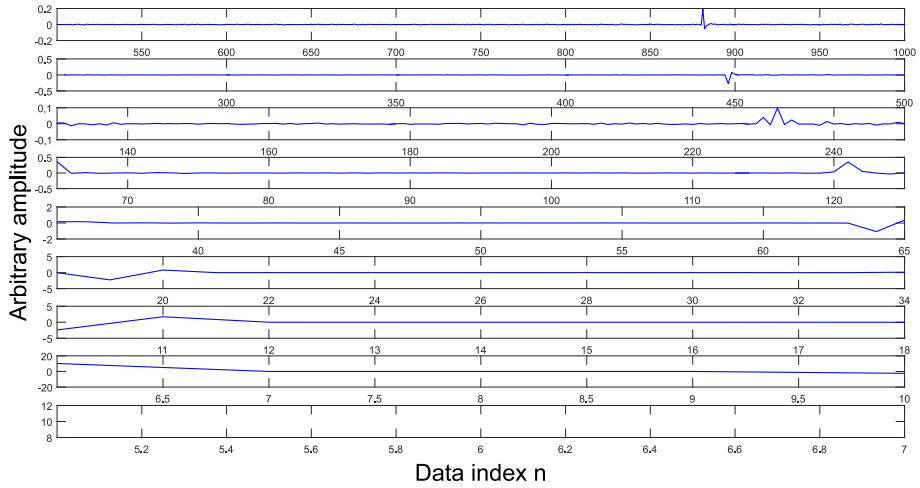


(c) Application of CWT to fault case (shearing tail)

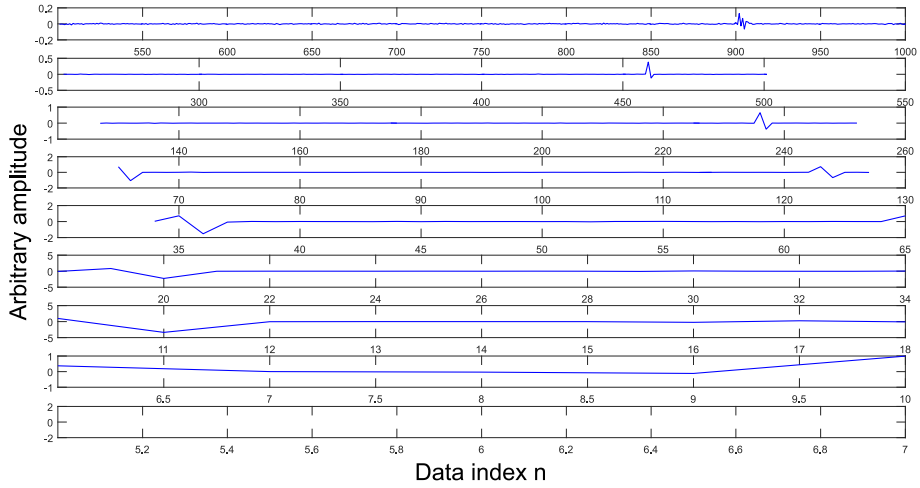
Figure 5.3: Graphical results of CWT



(a) Application of DWT to fault-free case [19]

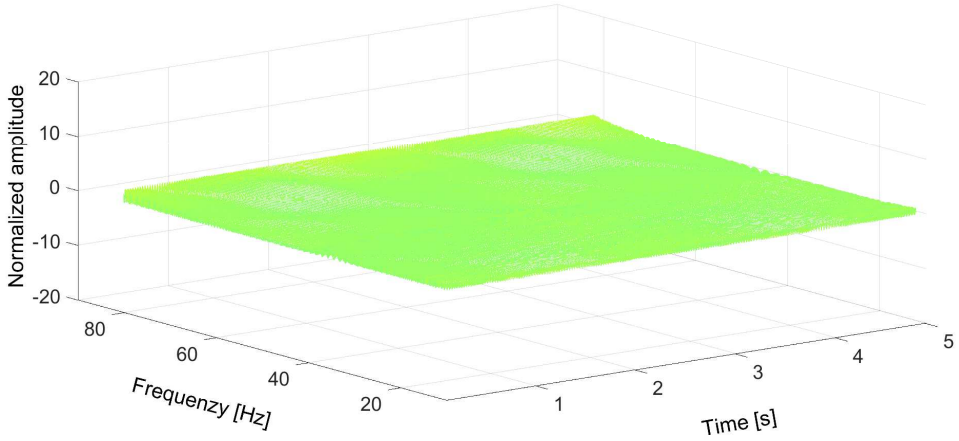


(b) Application of DWT to fault case (cobble) [19]

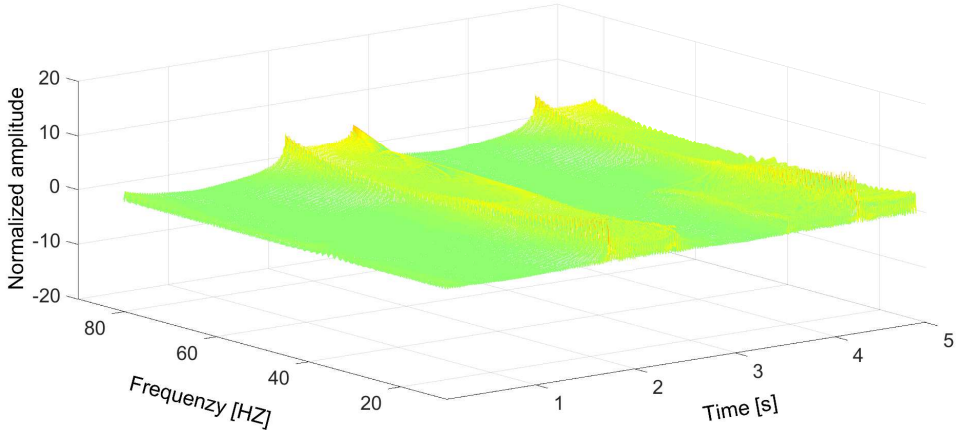


(c) Application of DWT to fault case (shearing tail)

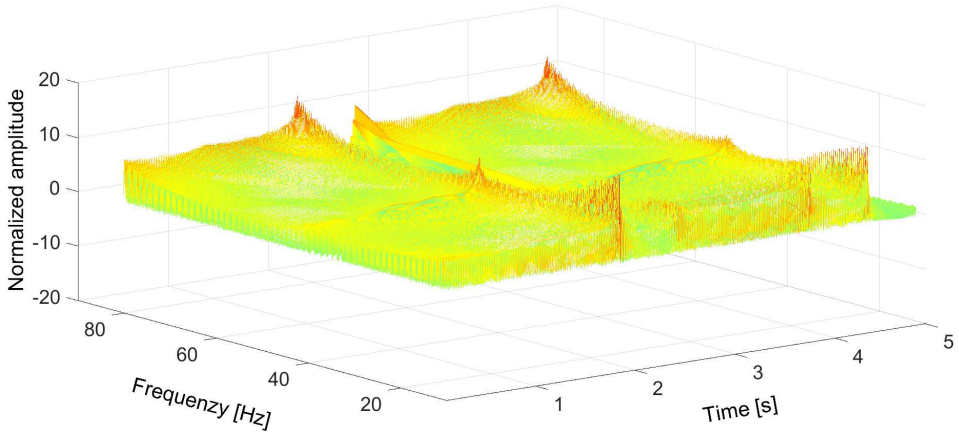
Figure 5.4: Graphical results of DWT



(a) Application of WVD to fault-free case [19]

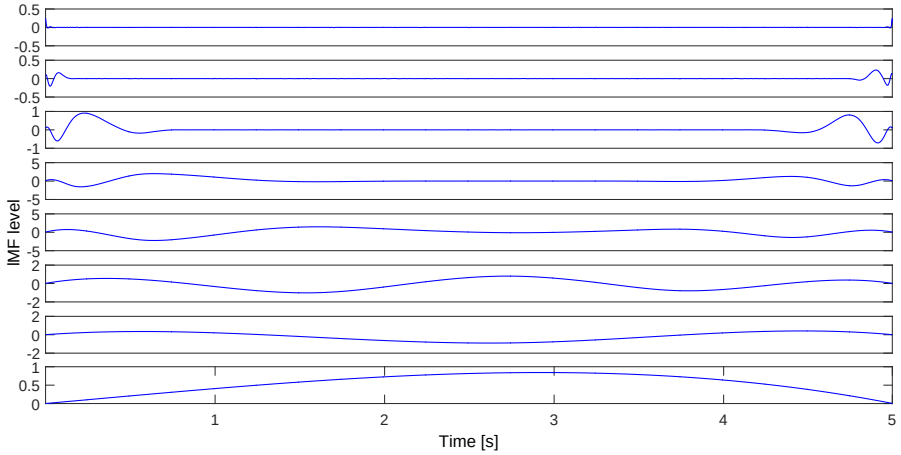


(b) Application of WVD to fault case (cobble) [19]

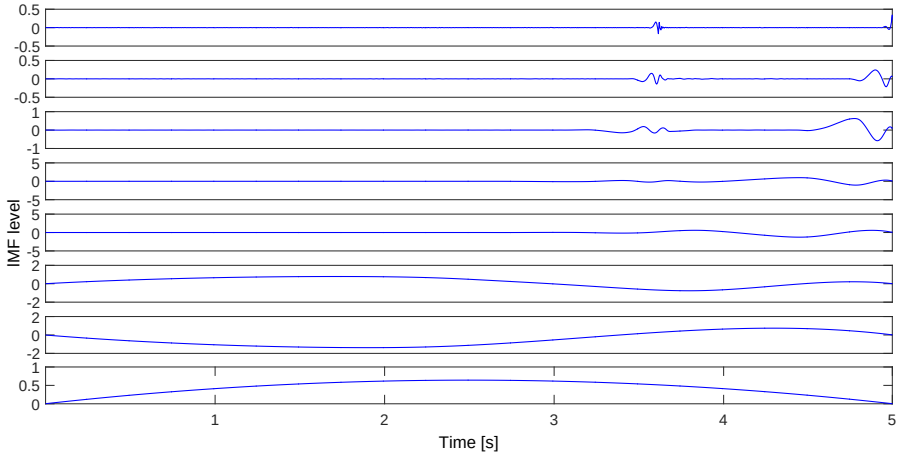


(c) Application of WVD to fault case (shearing tail)

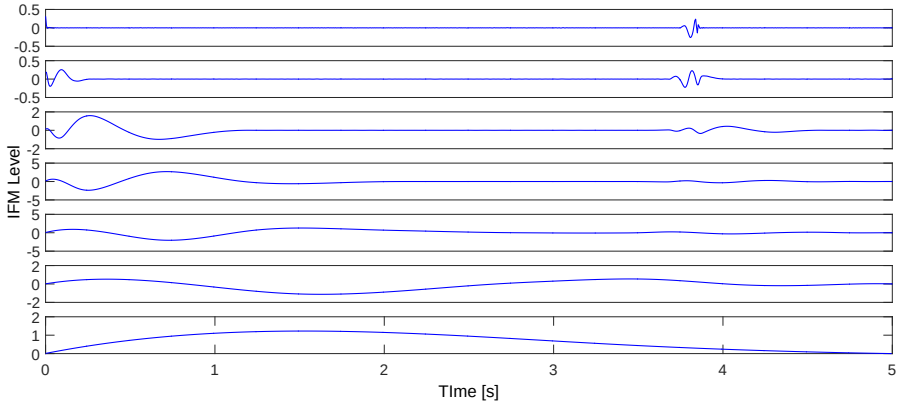
Figure 5.5: Graphical results of WVD



(a) Application of EMD to fault-free case [19]



(b) Application of EMD to fault case (cobble) [19]



(c) Application of EMD to fault case (shearing tail)

Figure 5.6: Graphical results of EMD

Summary of Properties

The properties of the presented methods are summarized in Table 5.1. Computational load and applicability are derived from the applied algorithms. All methods transform the time-based input signal into the time-frequency domain. The STFT preserves constant resolution for all frequencies, since the window is the same for the entire signal. Good frequency resolution at low frequencies (wide windows) comes with a low time resolution at high frequencies. The advantage of STFT is the easy interpretation of the result. That is why it is used in numerous investigations concerning acoustics and vibrations, where the square of the transform result is plotted as a spectrogram. At WT, the resolution can be adapted via the width of the window function. Therefore, it is well suited for non-stationary signals. The computational load and the amount of generated data are significant for CWT. Both can be reduced in DWT. The only non-linear transformation in this list is WVD. Due to interference terms, it is difficult to analyze the results if the original signal contains several frequency elements, which is the case here as it is in most practical applications. Commonly, EMD is applied together with HT to HHT. The advantages of this method are the high resolution and the good applicability to non-staionary signals, coming together with medium computational load.

Table 5.1: Properties of the presented methods [19]

	domain	resolution	computational load	linear	applicability to non-stationary signals
STFT	time-frequency	limited	low	yes	bad
CWT	time-frequency	variable	high	yes	good
DWT	time-frequency	variable	low	yes	good
WVD	time-frequency	high	medium	quadratic	satisfying
EMD	time-frequency	high	medium	yes	good

5.1.2 Classification results

The methods STFT, CWT, DWT, WVD, and EMD are applied to the input signals for fault detection. To classify the data, a SVM is trained with a selection from the feature extraction of the pre-processed input signals and tested with the remaining input signals. Alternatively, the classification is done in new way via symmetry degree of IMF's of given input signals. The application of the method combinations to the database broadened by cross-validation gives the results shown in Table 5.2. Correctly detected deviations in system behavior are displayed as true positive (TP), correct classification as regular system behavior are displayed as true negative (TN). In contrast, deviations in system behavior that are not detected are displayed as false negative (FN), and states falsely classified as deviation are displayed as false positive (FP).

In Table 5.2, the absolute number n of classified data sets is shown, followed by the relative number m, given as a percentage ($n : m$). Disregarding STFT-SVM, the results for TP and TN are lying between 85% and 100%. The performance of STFT-SVM is the worst with 15% TN and 50% TP. In contrast, the method combination of EMD-CC leads to the best results, showing 100% for TP and 95% for TN.

Disregarding STFT, a FP result appears for 5% to 10% of datasets and FN for 0% to 15%. Again, EMD-CC leads to the best results, zero cases for FN and two cases, respectively, 5% for FP. The performance of STFT-SVM is poor with 85% FP and 50% FN.

Of course, a high number of TP and a low number of FP is desirable. The case of FP is of great importance in practical applications, since a possibly fault-free process would be aborted inadvertently.

Table 5.2: Classification: detection rate

	STFT-SVM	CWT-SVM	DWT-SVM	WVD-SVM	EMD-SVM	EMD-CC
TP	20: 50%	36: 90%	34: 85%	37: 92,5%	38: 95%	40: 100%
TN	6: 15%	38: 95%	38: 95%	38: 95%	36: 90%	38: 95%
FP	34: 85%	2: 5%	2: 5%	2: 5%	4: 10%	2: 5%
FN	20: 50%	4: 10%	6: 15%	3: 7,5%	2: 5%	0: 0%

The receiver operating characteristic (ROC) gives graphically the performance of a binary classification. In a ROC space, the true positive rate (TPR) is plotted against the false positive rate (FPR). Each classification result of Table 5.2 is represented by one point in

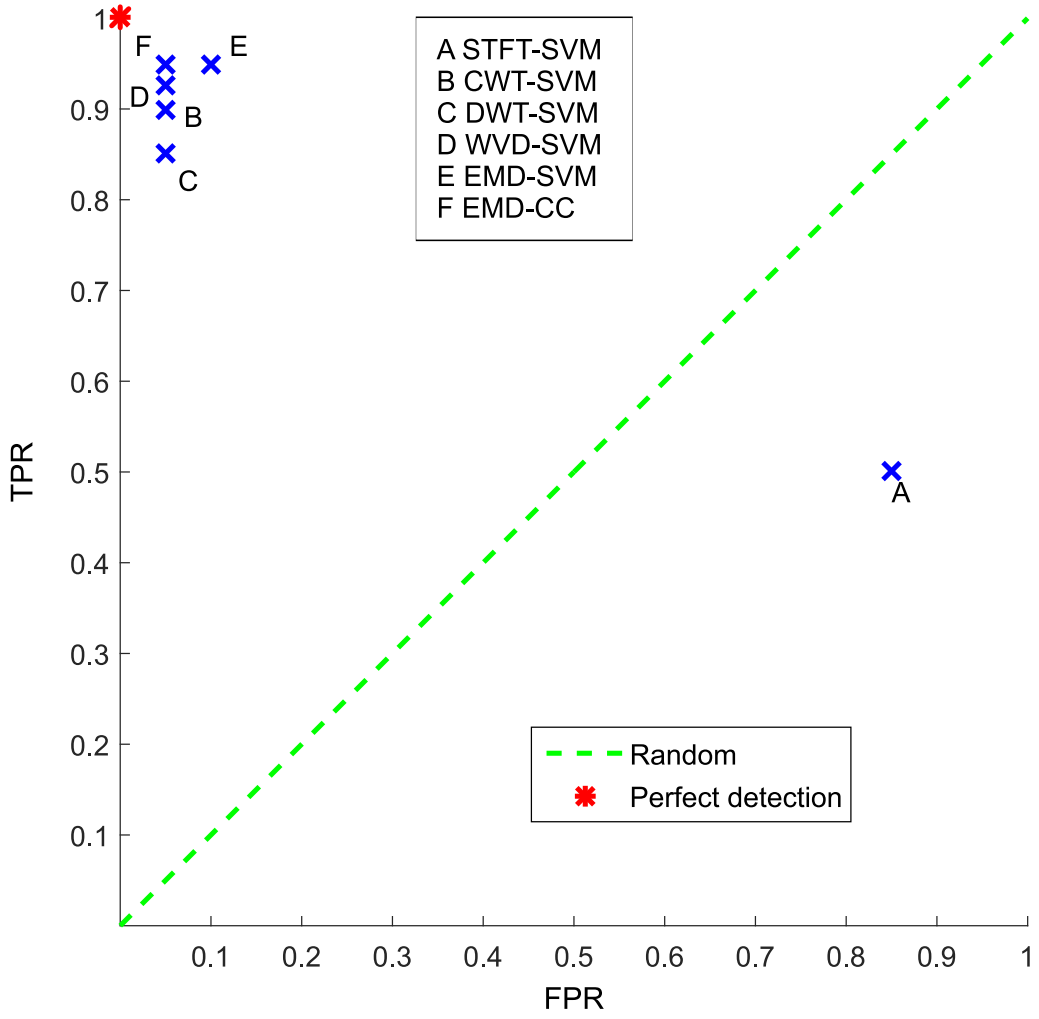


Figure 5.7: ROC space presenting the detection rate; STFT-SVM is indexed with A, CWT-SVM with B, DWT-SVM with C, WVD-SVM with D, EMD-SVM with E and EMD-CC with F

the graph. The best possible classification would have 100% TPR and 0% FPR, giving a point in the upper left corner, marked in red in Figure 5.7. In figure 5.7, the method combination STFT-SVM is indexed with A, CWT-SVM with B, DWT-SVM with C, WVD-SVM with D, EMD-SVM with E and EMD-CC with F in blue. The dotted green line represents a random distribution. All points above that dotted green line work out better than random. Those underneath work worse, meaning a misinterpretation during classification.

It is clearly visible that only STFT-SVM (index A) performs really badly. The points representing the other five methods lean towards perfect prediction. The EMD-CC (index F) is closest to the perfect prediction point and shows the largest distance to the line of random result, meaning this method combination has the best balance between TPR and classification error.

5.1.3 χ^2-test

The data basis, on which the hypotheses are applied, is resumed in Table 5.3. Compared are the detection results of the combined methods STFT-SVM, CWT-SVM, DWT-SVM, WVD-SVM, EMD-SVM, and EMD-CC. The results of the hypothesis test are given in Table 5.4. The null hypothesis $H_{0.1}$ assumes random performance. For all tested method combinations, the null hypothesis $H_{0.1}$ can be rejected significantly, as shown in the left column of Table 5.4. This means that non of the methods is performing randomly. This is also true for STFT and consistent with the previous interpretation of Figure 5.7. From the distance of point A to the dotted green line of random distribution in the ROC space (Fig. 5.6), it can be inferred that the classification is unlikely to perform randomly. The position of point A beneath that line shows that the classification is systematically wrong.

The second null hypothesis $H_{0.2}$ assumes an 80% hit rate for the classification. This hypothesis is rejected for all method combinations as the second column of Table 5.4 shows. This means that none of the methods has a hit rate of 80% but is performing statistically significantly better or worse.

The third null hypothesis $H_{0.3}$ assumes a hit rate of 90%. This hypothesis can be rejected for STFT-SVM and EMD-CC. The probability of DWT-SVM to perform 90% is 1. The other methods perform near to 90% so the null hypothesis $H_{0.3}$ could not be rejected, as shown in the third column of Table 5.4. The STFT-SVM performs worse that 90%. The null hypothesis $H_{0.3}$ is rejected for EMD-CC because this newly introduced algorithm performs better than 90%.

5.1.4 McNemar's test

Based on Table 5.2, the conservative assumption is made that detection of an error appears at the same particular cases for all methods. A significance level of $p = 0,05$ is defined as appropriate to reject the null hypothesis. In contrast to the standard hypothesis test, it is legit to draw conclusions from the rejection of the null hypothesis in a McNemar's test. The results of the application of a McNemar test are given in Table 5.5. It becomes clear that the STFT-SVM is performing vastly differently from the other method combinations. With regard to the numbers in Table 5.3, it is stated that STFT-SVM performs worse. The performance of CWT-SVM, DWT-SVM, WVD-SVM, and EMD-SVM is comparable, the null hypothesis can not be rejected. The performance of EMD-SVM and WVD-SVM is statistically not distinguishable. EMD-CC performs statistically different from STFT-SVM, and DWT-SVM, in this case better.

Table 5.3: Detection results for χ^2-test and McNemar's test

	STFT-SVM	CWT-SVM	DWT-SVM	WVD-SVM	EMD-SVM	EMD-CC	Random	80%	90%
True	26	74	72	75	74	78	40	64	72
False	54	6	8	5	6	2	40	16	8

Table 5.4: Values of χ^2-test applied to detection results

	Random	80%	90%
STFT-SVM	$\chi^2 = 9.8$ $\rightarrow p << 0.01$	$\chi^2 = 112.8$ $\rightarrow p << 0.001$	$\chi^2 = 293.9$ $\rightarrow p << 0.001$
CWT-SVM	$\chi^2 = 57.8$ $\rightarrow p << 0.001$	$\chi^2 = 7.81$ $\rightarrow p < 0.01$	$\chi^2 = 0.56$ $\rightarrow p \approx 0.4$
DWT-SVM	$\chi^2 = 51.2$ $\rightarrow p << 0.001$	$\chi^2 = 5$ $\rightarrow p \approx 0.025$	$\chi^2 = 0$ $\chi^2 =\rightarrow p = 1$
WVD-SVM	$\chi^2 = 61.25$ $\rightarrow p << 0.001$	$\chi^2 = 9.45$ $\rightarrow p < 0.01$	$\chi^2 = 1.25$ $\rightarrow p \approx 0.3$
EMD-SVM	$\chi^2 = 57.8$ $\rightarrow p << 0.001$	$\chi^2 = 7.81$ $\rightarrow p < 0.01$	$\chi^2 = 0.56$ $\rightarrow p \approx 0.4$
EMD-CC	$\chi^2 = 72.2$ $\rightarrow p << 0.001$	$\chi^2 = 15.31$ $\rightarrow p < 0.001$	$\chi^2 = 5$ $\rightarrow p \approx 0.025$

5.2 Change identification

The fault detection discussed in Chapter 5.1 is of prime importance in practical applications, since the reaction at the production site will be similar for both faults discussed here. Fundamentally, all four states regarded here have to be identified. The evaluation for fault identification has to differentiate four system states from each other. Therefore, the detection rate of fault identification is regarded, meaning the probability to distinguish between four system states. Regarding all four system states separately, TP, TN, FP, and FN are redefined. The true classification of a particular state is marked as TP, the correct negation as TN. A missed identification of the state is listed as FN. A state erroneously classified as this specific state is listed as FP, false positive.

5.2.1 Classification results

Table 5.6 shows the results for FP and TP classification. The total number of data stets for each state is eighty in the data set broadened by cross-validation. Shown in the table are the absolute and the relative values for all states and all methods. The relative values are calculated with respect to a total number of twenty for TP and a total number of sixty for FP. The absolute number n of classified data sets is followed by the relative number m, given as a percentage $(n : m)$. The States 1-4 are defied in Section 4.1.

No method is able to identify all states. Again, EMD-CC shows the best over-all results. For EMD-CC, all TP rates of fault identification lie between 80% and 90%. The fault State 3 is identified in 80% and the fault State 4 in 90% of the cases. Second best performing is DWT-SVM, giving values between 75% and 100%. Fault State 3 is identified in 75% and fault State 4 in 80% of the cases. Both methods are able to differentiate between all four states. For the fault states, the FP identification is 3.3% respectively 6.7% with EMD-CC and 3.3% respectively 5% with DWT-SVM. The detailed results show that in most FP cases State 3 and State 4 are mixed up. This means that the deviation is correct detected but identified wrongly.

Since the initial reaction of machine operators is the same in both fault states, this mixing-up of the states is of minor importance in practice.

The classification results of the state identification given in Table 5.6 are plotted as a ROC space. The States 1 and 2 are shown in Figure 5.8, State 3 and 4 in Figure 5.9. In this case, the TPR is plotted against the FPR. In both figures, the best possible classification with 100% identification and 0% FPR is given in a red point in the upper left corner. The dotted green line represents a random distribution. The method combination STFT-SVM is indexed with A, CWT-SVM with B, DWT-SVM with C, WVD-SVM with D, EMD-SVM with E and EMD-CC with F.

In Figure 5.8, State 1 has the additional index 1, the data points are colored magenta. The index 2 is added for State 2, the data points are in cyan. In Figure 5.9, State 3 has the additional index 3, the data points are plotted in blue. Index 4 is added for State 4, data points are marked in green. All points above the dotted green line work out better than random. Those underneath the dotted linework worse, meaning a misinterpretation during classification. This representation points out which method combination is most suitable for each state, giving base for a possible decision fusion. Figure 5.8 visualizes that State 1 is best classified by EMD-CC (F1). State 2 is best classified by DWT (C2) and similarly good by EMD-CC (F2). Figure 5.9 visualizes that State 3 and State 4 are best classified by EMD-CC (F3, F4). Both figures show that STFT-SVM is not suited. The points with index A for STFT are lying near the line of random result.

Table 5.5: McNemar's test

	CWT-SVM	DWT-SVM	WVD-SVM	EMD-SVM	EMD-CC
STFT-SVM	$\chi^2 = 46,02$ $\rightarrow p << 0.001$	$\chi^2 = 44.02$ $\rightarrow p << 0.001$	$\chi^2 = 47.02$ $\rightarrow p << 0.001$	$\chi^2 = 46.02$ $\rightarrow p << 0.001$	$\chi^2 = 50.02$ $\rightarrow p << 0.001$
CWT-SVM		$\chi^2 = 0.5$ $\rightarrow p < 0.5$	$\chi^2 = 1.33$ $\rightarrow p < 0.3$	$\chi^2 = 1$ $\rightarrow p < 0.4$	$\chi^2 = 2.25$ $\rightarrow p < 0.2$
DWT-SVM	$\chi^2 = 0.5$ $\rightarrow p < 0.5$		$\chi^2 = 1.33$ $\rightarrow p < 0.3$	$\chi^2 = 0.5$ $\rightarrow p < 0.5$	$\chi^2 = 4.17$ $\rightarrow p < 0.05$
WVD-SVM	$\chi^2 = 1.33$ $\rightarrow p < 0.3$	$\chi^2 = 1.33$ $\rightarrow p < 0.3$		$\chi^2 = 0$ $\rightarrow p = 1$	$\chi^2 = 1.33$ $\rightarrow p < 0.3$
EMD-SVM	$\chi^2 = 1$ $\rightarrow p < 0.4$	$\chi^2 = 0.5$ $\rightarrow p < 0.5$	$\chi^2 = 0$ $\rightarrow p = 1$		$\chi^2 = 2.25$ $\rightarrow p < 0.2$

Table 5.6: Classification: Detection rate of fault identification

	STFT-SVM	CWT-SVM	DWT-SVM	WVD-SVM	EMD-SVM	EMD-CC
State 1						
TP	3: 15%	15: 75%	18: 90%	17: 85%	20: 100%	18: 90%
TN	60: 100%	55: 91.7%	60: 100%	57: 93.3%	50: 83.3%	58: 96.7%
FP	0: 0%	5: 8.3%	0: 0%	3: 5%	10: 16.7%	2: 3.3%
FN	17: 85%	5: 25%	2: 10%	3: 15%	0: 0%	2: 10%
State 2						
TP	3: 15%	14: 70%	20: 100%	14: 70%	8: 40%	18: 90%
TN	55: 91.7%	52: 86.7%	54: 90%	53: 88.3%	55: 91.7%	56: 93.3%
FP	5: 8.3%	8: 13.3%	6: 10%	7: 11.7%	5: 8.3%	4: 6.7%
FN	17: 85%	6: 30%	0: 0%	6: 30%	12: 60%	2: 10%
State 3						
TP	0: 0%	12: 60%	15: 75%	11: 55%	6: 30%	16: 80%
TN	56: 93.3%	53: 88.3%	58: 96.7%	53: 88.3%	55: 91.7%	58: 96.7%
FP	4: 6.7%	7: 11.7%	2: 3.3%	7: 11.7%	5: 8.3%	2: 3.3%
FN	20: 100%	8: 40%	5: 25%	9: 45%	14: 70%	4: 20%
State 4						
TP	19: 95%	15: 75%	16: 80%	16: 80%	18: 90%	18: 90%
TN	14: 23.3%	56: 93.3%	57: 95%	55: 91.7%	48: 80%	56: 93.3%
FP	46: 76.7%	4: 6.7%	3: 5%	5: 8.3%	12: 20%	4: 6.7%
FN	1: 5%	5: 25%	4: 20%	4: 20%	2: 10%	2: 10%

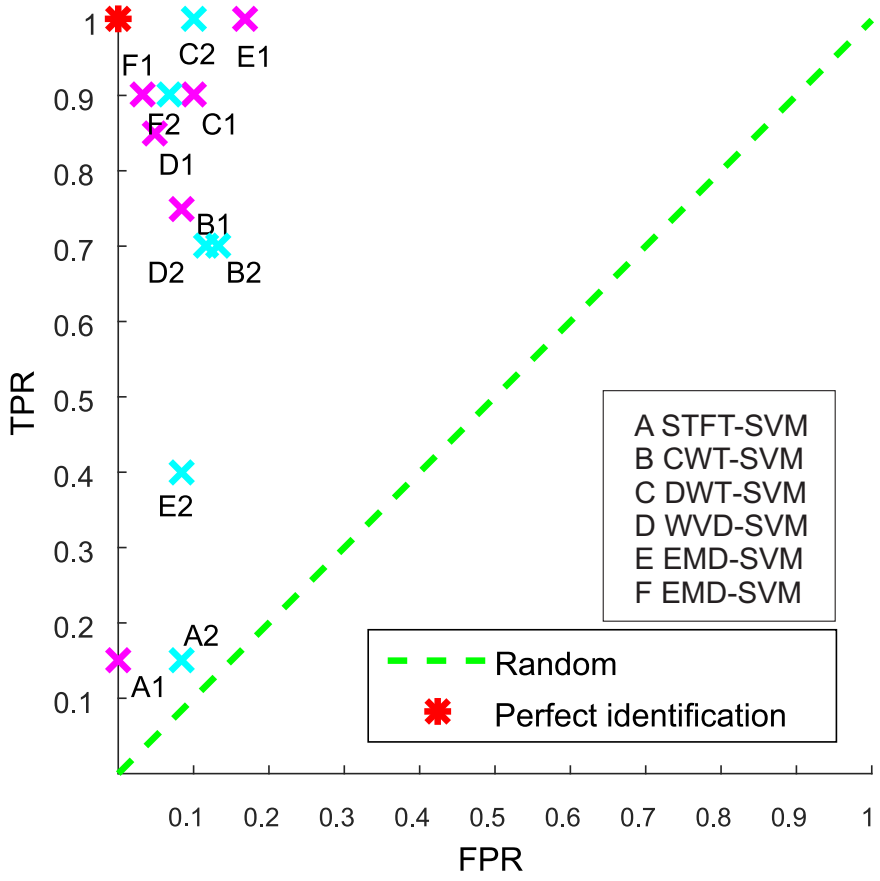


Figure 5.8: ROC space presenting detection rate of fault identification of State 1 and State 2; STFT-SVM is indexed with A, CWT-SVM with B, DWT-SVM with C, WVD-SVM with D, EMD-SVM with E and EMD-CC with F

5.2.2 χ^2-test

Likewise as in Section 5.1.3, a hypothesis test is performed with equivalent constraints. Again, the first null hypothesis $H_{0.1}$ is assuming a normal distribution or random occurrence for the results. If the null hypothesis cannot be rejected, this is not sufficiently significant to accept the null hypothesis, but the method is not proven to be better than randomly chosen results. The second null hypothesis $H_{0.2}$ is assuming that the results of the respective method will hit the correct class with a probability of 80%. The value of 80% is chosen corresponding to the observed results. If the compared method performs better or worse, the null hypothesis will be rejected. A significance level of $p = 0,05$ is defined as appropriate to reject the null hypothesis.

Table 5.7 resumes the data base of the detection rates of fault identification of STFT-SVM, CWT-SVM, DWT-SVM, WVD-SVM, EMD-SVM, and EMD-CC. Table 5.8 lists the results of the χ^2-test. For all method combinations, the null hypothesis $H_{0.1}$ can be

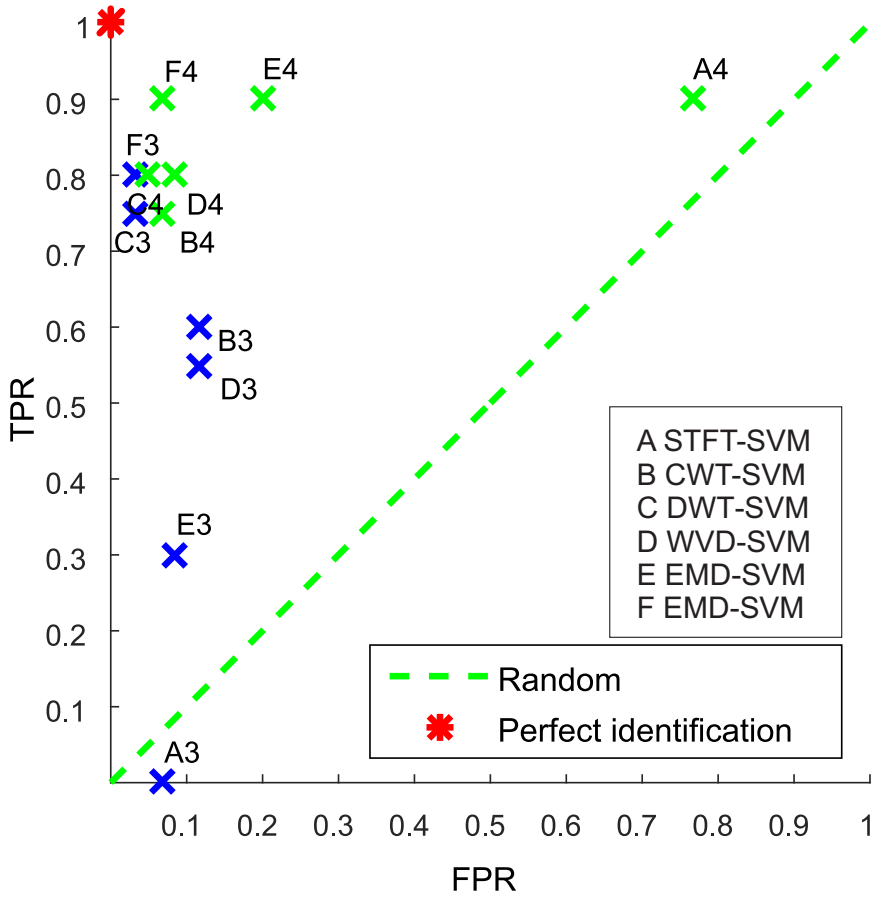


Figure 5.9: ROC space presenting detection rate of fault identification of State 3 and 4; STFT-SVM is indexed with A, CWT-SVM with B, DWT-SVM with C, WVD-SVM with D, EMD-SVM with E and EMD-CC with F

rejected. None of them is performing randomly. The second null hypothesis $H_{0.2}$ can be significantly rejected for STFT-SVM, CWT-SVM, and EMD-SVM. Lokking at the absolute values in Table 5.7, it becomes clear that these methods are performing worse than 80%. The performance of the other method combinations is statistically similar to 80%. For these, the null hypothesis $H_{0.2}$ could not be rejected.

Table 5.7: Detection rate of fault identification for χ^2-test and McNemar's test

	STFT-SVM	CWT-SVM	DWT-SVM	WVD-SVM	EMD-SVM	EMD-CC	Random	80%
True	25	56	64	58	52	70	40	64
False	55	24	16	22	28	10	40	16

Table 5.8: Results of χ^2-test applied to detection rate of fault identification

	Random	80%
STFT-SVM	$\chi^2 = 11,25 \rightarrow p \approx 0.001$	$\chi^2 = 118.83 \rightarrow p << 0.001$
CWT-SVM	$\chi^2 = 12.8 \rightarrow p < 0.001$	$\chi^2 = 5 \rightarrow p \approx 0.025$
DWT-SVM	$\chi^2 = 28.8 \rightarrow p < 0.001$	$\chi^2 = 0 \rightarrow p \approx 1$
WVD-SVM	$\chi^2 = 16.2 \rightarrow p < 0.001$	$\chi^2 = 2.81 \rightarrow p < 0.1$
EMD-SVM	$\chi^2 = 7.2 \rightarrow p < 0.01$	$\chi^2 = 11.25 \rightarrow p < 0.001$
EMD-CC	$\chi^2 = 45 \rightarrow p < 0.001$	$\chi^2 = 2.81 \rightarrow p < 0.1$

5.2.3 McNemar's test

A significance level of $p = 0.5$ is defined as appropriate to reject the null hypothesis that the methods are performing similarly. Compared are STFT-SVM to CWT-SVM to DWT-SVM to WVD-SVM to EMD-SVM to EMD-CC. Details on the results are given in Table 5.9. The probability level is given in percent. The rejection of the null hypothesis does not give any information about the quality of the compared methods. This information has to be derived from the results themselves. The results in Table 5.9 show that STFT-SVM is performing significantly differently from the other method combinations. Table 5.7 completes this information and shows that it is performing significantly worse. The WVD-SVM and EMD-SVM perform statistically similar to CWT-SVM. Best performing is EMD-CC: statistically significantly better than the second best performing method combination DWT-SVM and way better than the other compared method combinations. As discussed above, these are the both methods able to distinguish the four states.

Table 5.9: McNemar's test

	CWT-SVM	DWT-SVM	WVD-SVM	EMD-SVM	EMD-CC
STFT-SVM	$\chi^2 = 29.03$ $\rightarrow p << 0.001$	$\chi^2 = 37.02$ $\rightarrow p << 0.001$	$\chi^2 = 31.03$ $\rightarrow p << 0.001$	$\chi^2 = 25.03$ $\rightarrow p << 0.001$	$\chi^2 = 43.02$ $\rightarrow p << 0.001$
CWT-SVM		$\chi^2 = 6.12$ $\rightarrow p \approx 0.01$	$\chi^2 = 0.5$ $\rightarrow p \approx 0.5$	$\chi^2 = 2.25$ $\rightarrow p < 0.2$	$\chi^2 = 12.07$ $\rightarrow p \approx 0.001$
DWT-SVM	$\chi^2 = 6.12$ $\rightarrow p \approx 0.01$		$\chi^2 = 4.16$ $\rightarrow p < 0.05$	$\chi^2 = 10.08$ $\rightarrow p \approx 0.001$	$\chi^2 = 4.16$ $\rightarrow p < 0.05$
WVD-SVM	$\chi^2 = 0.5$ $\rightarrow p \approx 0.5$	$\chi^2 = 4.16$ $\rightarrow p < 0.05$		$\chi^2 = 4.16$ $\rightarrow p < 0.05$	$\chi^2 = 10.08$ $\rightarrow p \approx 0.001$
EMD-SVM	$\chi^2 = 2.25$ $p < 0.2$	$\chi^2 = 10.08$ $\rightarrow p \approx 0.001$	$\chi^2 = 4.16$ $\rightarrow p < 0.05$		$\chi^2 = 16.05$ $\rightarrow p < 0.001$

5.3 Fault prognosis

For severe faults, prediction is highly desirable to avoid system damage and long down-times. The prediction method has to be precise enough to avoid false alarms that lead to loss of production time on the one hand, and on the other hand not to miss upcoming faults to reach the desired detection. In this work, the best performing signal analysis method for the tested set-up is used: EMD-CC. Additionally, to the good classification performance, a benefit of EMD-CC is the relatively low computational load.

The method's suitability for fault prediction from one strip to an other is tested. The special case of two consecutive strips of the same grade with the same dimension was analyzed for strip-to-strip prediction. Here, the first strip passed without disturbances, the second caused a cobble. Even in that case, no critical deviations in State 1 and State 2 could be detected.

Rolling is a dynamic process and faults treated in this approach occur rapidly. For this reason, the prediction test is applied to signal segments in immediate advance of the fault.

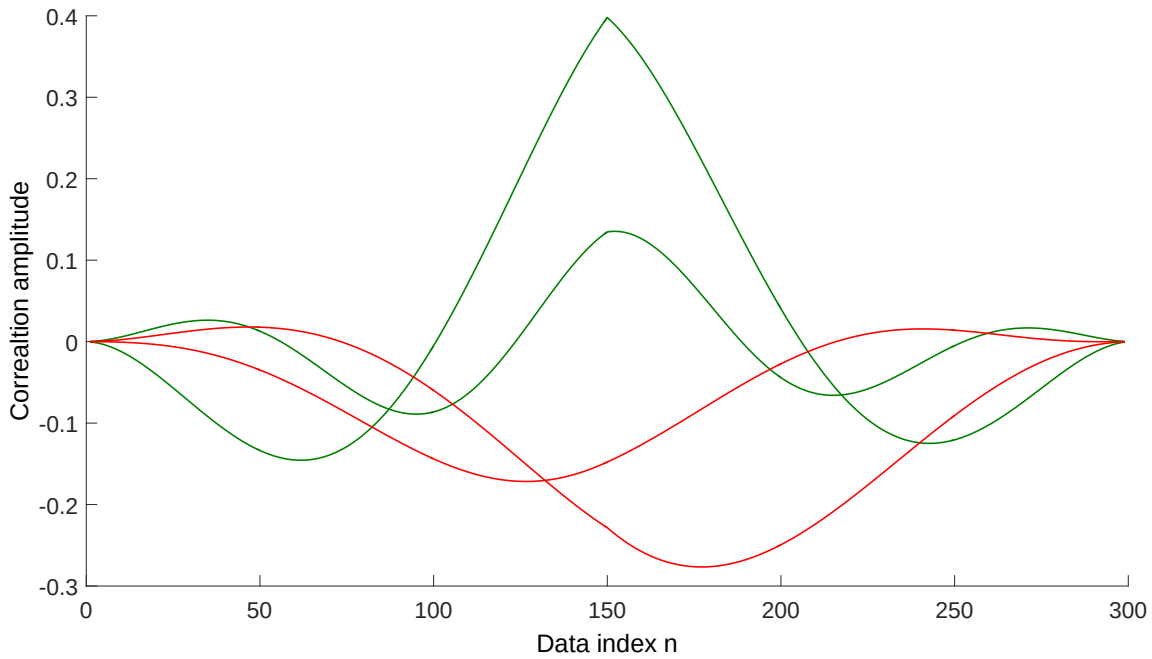


Figure 5.10: Prognosis EMD-CC; green line: State 1, red line: State 3 [179]

5.3.1 Classification results

From the event-based data, the fault occurrence time is determined. Signal segments in advance of that time mark are generated and EMD-CC is applied. Similar features as

described in Section 5.1 and 5.2 are visible. The amplitude and the maxima of the IMF coefficients give information on the system state. Figure 5.10 shows exemplary correlation-coefficients of IMF 5 and 6. Both dotted green lines represent a regular system state (State 1), and the red lines represent a signal captured two seconds before the occurrence of a cobble (State 3).

In practice, the distinction between regular system behavior and deviated system behavior is of importance. Therefore, the fault prediction is regarded as a fault detection task, first. Table 5.10 gives the details. Truly detected deviations in system behavior are displayed as true positive (TP), true classification as regular system behavior are displayed as true negative (TN). In contrast, deviations in system behavior that are not detected are displayed as false negative (FN), and falsely as deviation classified states are displayed as false positive (FP). The absolute number n of classified data sets is followed by the relative number m, given in percentage ($n : m$).

In the prediction of faults, the EMD-CC achieves 100% TP with 7.5% FP. This means, that all faults are detected, but three of the fault-free samples are indexed as faults.

Table 5.10: Classification: Detection rate of fault prediction

	EMD-CC
TP	40: 100%
TN	37: 92.5%
FP	3: 7.5%
FN	0: 0%

The ROC space plot of the results in Table 5.10 is given in Figure 5.11. The TPR is plotted against the FPR, the best possible classification is marked in red and the random distribution is represented by the dotted green line.

It is clearly visible that the prognosis is far away from the random distribution, leaning towards the point of perfect prediction.

The detection of the two specified faults is important in practical applications. Additionally, the performance of the evaluated methods concerning the identification of all four system states is complemented. The results of the identification are listed in Table 5.11. The TPR is between 95% and 65%, State 1 is identified best. The FPR is with 0% lowest for State 2, and lies between 8.3% and 11,7% for the three other states.

Table 5.11: Classification: Detection rate of fault identification prediction with EMD-CC

	State 1	State 2	State 3	State 4
TP	19: 95%	13: 65%	13: 65%	17: 85%
TN	55: 91.7%	60: 100%	54: 90%	53: 88.3%
FP	5: 8.3%	0: 0%	6: 10%	7: 11.7%
FN	1: 5%	7: 35%	7: 35%	3: 15%

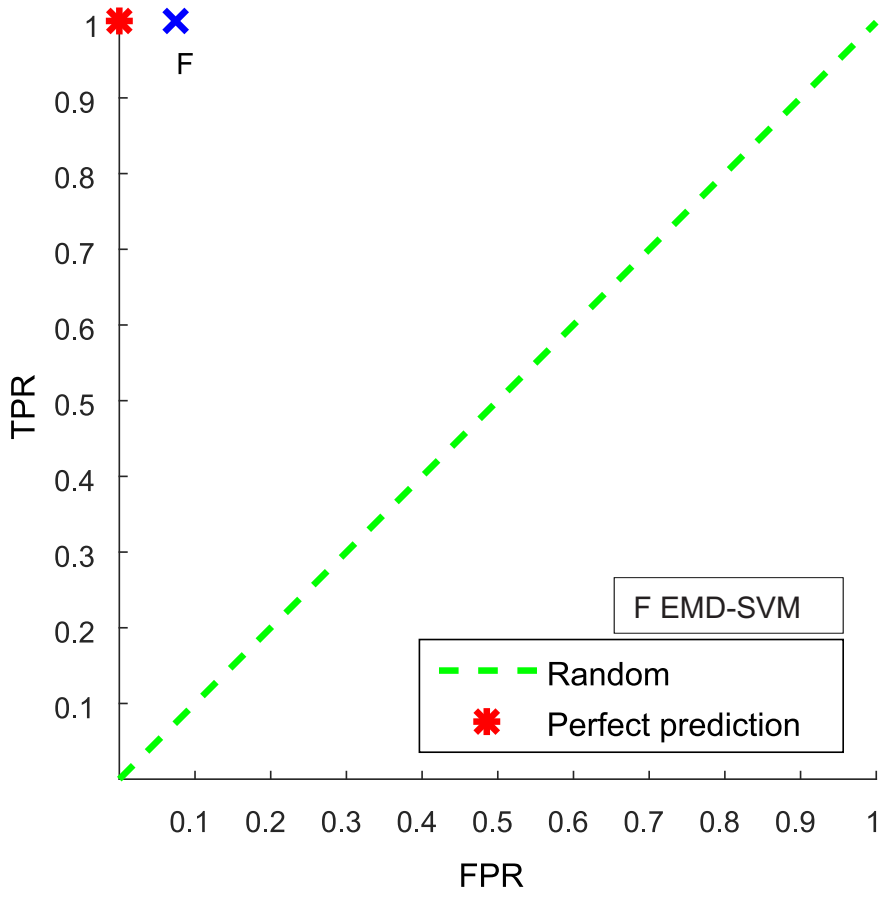


Figure 5.11: ROC space presenting the fault detection rate in matters of prognosis; EMD-CC is indexed with F

The results shown in Table 5.11 are plotted in a ROC space graph in Figure 5.12. The TPR is plotted against the FPR, the best possible classification is marked in red and the random distribution is represented by the dotted green line. State 1 has the additional index 1, the data point is colored magenta. The index 2 is added for State 2, the data point is in cyan. State 3 has the additional index 3, the data point is plotted in blue. Index 4 is added for State 4, data point is marked in green. Figure 5.12 visualizes that State 1 is identified best and State 3 worst.

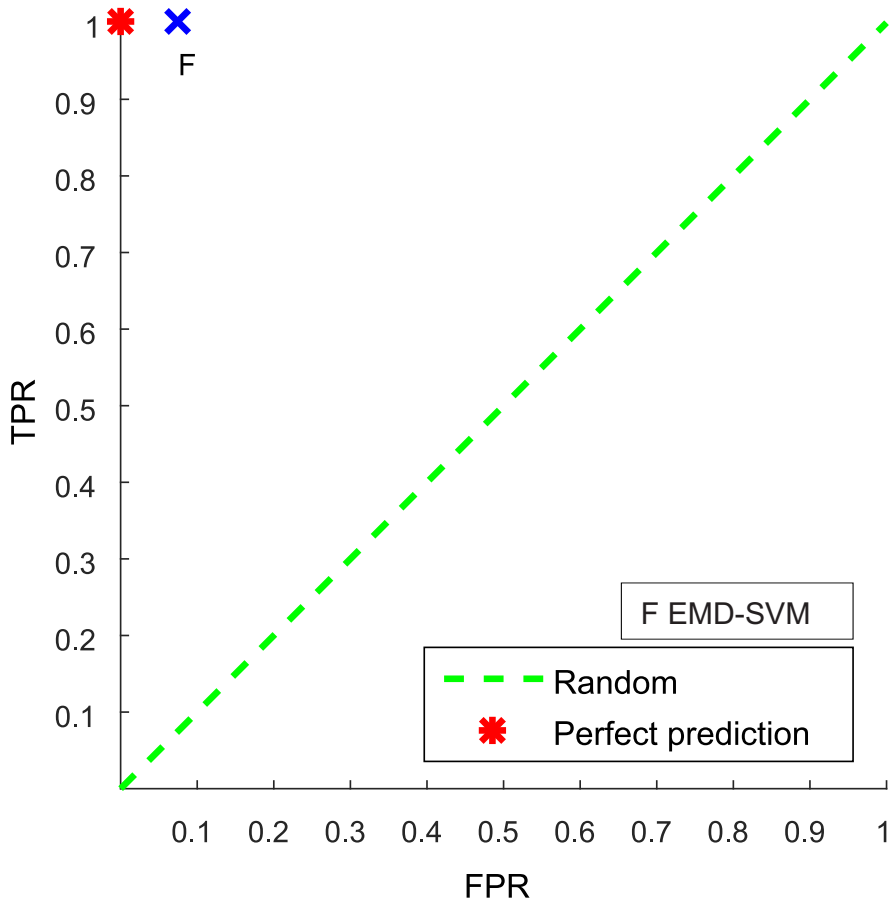


Figure 5.12: ROC space presenting the detection rate of fault identification of State 1-4 in matters of prognosis; EMD-CC is indexed with F

5.3.2 χ^2-test

A hypothesis test is performed on the results of the prognosis in case of detection as well as of identification, given in Table 5.12. Again, the first null hypothesis $H_{0.1}$ is assuming a normal distribution or random occurrence for the results. The second null hypothesis $H_{0.2}$ is assuming that the results of the respective method will hit the correct class with a probability of 80%. The value of 80% is chosen corresponding to the observed results. If the compared method performs better or worse, the null hypothesis will be rejected. A significance level of $p = 0,05$ is defined as appropriate to reject the null hypothesis.

The results are given in Table 5.13. The null hypothesis $H_{0.1}$ can be significantly rejected for prognosis detection as well as prognosis identification. Both approaches perform better than random. The second null hypothesis $H_{0.2}$ can be rejected significantly for prognosis detection, whereas prognosis identification performs statistically similar to 80% TPR. The null hypothesis $H_{0.2}$ cannot be rejected.

Table 5.12: Classification rate for χ^2-test

	Detection EMD-CC	Identification EMD-CC	Random	80%
True	77	62	40	64
False	3	18	40	16

Table 5.13: Results of χ^2-test applied to prediction rate

	Random	80%
Detection EMD-CC	$\chi^2 = 68.45$ $\rightarrow p << 0,001$	$\chi^2 = 6.75$ $\rightarrow p < 0,01$
Identification EMD-CC	$\chi^2 = 24.2$ $\rightarrow p << 0,001$	$\chi^2 \approx 0$ $\rightarrow p \approx 100\%$

5.4 Discussion

The results of the application of all five methods to real data of a hot strip rolling mill reveal that CWT and WVD come along with practical problems. The computation time for CWT is considerably higher than that for the other methods. The number of data points scales quadratically, so for a typical data set length of a thousand the number of points in the result is one million.

The number of data points in the result of WVD is also squared, whereas the calculation time is comparable to the other methods. The WVD's main shortcoming is the appearance of interference terms due to the quadratic behavior of the method.

The results of the presented STFT application show a strong effect of the fault in the time scale. A broad distribution of energy along the frequency axis can be seen. The fault can be localized, but its identification is difficult, not to say impossible. So, for faults like cobble and shearing tail, the method does not seem suitable. The application of CWT does not resolve the timing of the fault. Instead, strong changes in the result appear several seconds before the fault occurs. Since other methods perform better in detection and identification of the faults, CWT has not been considered for prognosis. In contrast to CWT with 3D graphics of the results, DWT gives data vectors in several decomposition levels. The DWT method is the one with the lowest computation time and the smallest number of data points in the result, namely the same number as in the original data

set. The fault occurrence is precisely indicated by a glitch in the high frequency part of the result. Further evaluation of the resulting data vectors needs expert knowledge. The results of WVD show too many interference terms due to the multi-frequency terms of real data. Filtering these interference terms needs pre-knowledge of the interference frequency bands, therefore it is not suitable for these applications. The EMD splits a signal into several vectors. As with DWT, the interpretation of EMD needs expert knowledge. For each IMF, the number of data points remains constant, so that the total number typically is seven times higher than that of the original data. An automated interpretation of the EMD results is possible by additional mathematical treatment for classification.

One of the applied classification approaches is a SVM. The performance of a classifier is strongly depending on the quality of feature extraction. This is shown by the results in Table 5.2 and Table 5.6. Still, four methods performed better than random in case of detection and in case of identification. Only STFT showed a consequent miss-interpretation of the features. In case of detection, EMD-SVM performed better than the other SVM combinations. In case of identification, DWT had the best over all performance. In detail, State 1 was best detected by EMD-SVM, State 2 by DWT-SVM, State 3 by DWT-SVM, and State 4 by EMD-SVM. As a new classification function, the EMD is supplemented by CC. The results are thresholded for interpretation. This method combination outperformes all the SVM approaches in matters of detection and identification of all four states.

To check the probability of fault prediction, the best performing method EMD-CC is used. The prognosis from strip to strip is not possible. Regarding time slices in immediate advance, EMD-CC is able to detect all faults with a FP rate of 7.5%. In case of identification, State 1 and 4 are classified best. In a practical application, the distinction between the upcoming faults is not a prior task, the detection is sufficient. Therefore, fault prediction and generation of an alarm seems possible with the new method of EMD-CC.

6 Summary and future work

6.1 Summary

The presented work investigates the condition monitoring of the complex production process of a hot strip rolling mill. A signal-based fault diagnosis and fault prognosis approach for strip travel is developed. The new approach introduced here is able to detect two specific severe faults, to identify them, to distinguish between four different system states, and to give a prognosis on the system behavior.

In the first chapter of this work, the motivation for the investigation is given. A general description of the problems treated here and an overview on the methods tested in the present work for their suitability is presented.

In Chapter 2, a literature review gives an overview about previous research on related topics. It is shown that a great amount of the applications of fault diagnosis methods deal with rotating machinery and focus on periodic or quasi periodic signals. A small number of previous publications on the application of signal-based analysis in the field of strip rolling mills could be found. Only three of the five methods applied in the presented contribution, namely short time Fourier transform, discrete wavelet transform, and empirical mode decomposition, have been used in the previous applications. Continuous wavelet transform and Wigner Ville distribution, have not yet been applied to strip travel in strip rolling mills. None of the previous works solves the diagnosis problem of cobbles and shearing tails. Therefore, further investigation is necessary to provide a satisfactory solution.

In Chapter 3, the application site is introduced and the background of rolling is summed up. A brief glimpse is thrown on the forming process. The technical fundamentals of the strip rolling mill that provided the data used in this work are presented together with the fundamentals of rolling. Additionally, the target deviations in strip travel are described.

In Chapter 4, the design of the new signal processing chain is presented. Starting with the definition of system states, the selection of input signals, and the generation of data sets the signal processing steps are detailed. The selection of a suited input signal is an essential step to explore distinguishable features of different system states. The mathematical background on the pre-processing, the feature extraction, and the classification is

rolled out. The classification task is differentiated into fault detection, fault identification and fault prognosis. The proposed approach combines four different methods for feature extraction with two different classification algorithms. Combinations of these feature extraction and classification methods are applied to rolling force data originating from a hot strip mill. Especially, the suitability of the methods not yet applied to hot strip mills is evaluated. In this work, the new combination of empirical mode decomposition and cross-correlation is developed to make in-time fault diagnosis possible.

In Chapter 5, the results of the application to industrial data and their statistical validation is given. The applied signal-based methods perform differently in fault detection and fault identification. The occurrence time of faults is clearly indicated by short time Fourier transform, discrete wavelet transform, and empirical mode decomposition. Disregarding short time Fourier transform, the methods combined with support vector machine, respectively, cross-correlation are able to detect the two fault types treated in this work. The short time Fourier transform results show a misinterpretation of the features.

The performance in fault identification differs for the discussed methods. Again, short time Fourier transform combined with support vector machine is not able to identify the faults. Best results are achieved by empirical mode decomposition combined with cross-correlation. The new combination of empirical mode decomposition combined with cross-correlation has been used for fault prognosis. Usable information can be extracted in a time slice a few seconds in advance of the fault. With this information, empirical mode decomposition combined with cross-correlation is able to predict upcoming faults, and an alarm signal for machine operators can be generated.

6.2 Future work

The implementation of the presented fault detection approach into the production plant is proposed to solve specific condition monitoring tasks in hot strip mills. An alarm for machine operators can be generated or an automatized reaction executed. Once implemented, multiple other fault types can be treated and the fault detection performance in matters of other faults can be evaluated.

The available amount of data for this work was limited. The measured system data are stored for short-time only, because of limited disc space. Therefore, only faults occurred since the beginning of the research could be evaluated. Since this is an industrial production process and not a laboratory experiment, the faults cannot be provoked, but have to occur during the process. A linkage with the event-based data server might allow an automatized storage in case of certain faults. This way, the needed test and training data

for further studies of other fault types can be generated.

Another aspect to be realized in future work is to integrate a decision fusion approach. The introduced approaches have individual performance rates for the four different system states. In case of uncertain classifications, an additional feature extraction and classification step might improve the false positive rate.

Bibliography

[1] R. Isermann, Fault-Diagnosis Systems, Springer Verlag, Berlin Heidelberg, 2006.

[2] R. J. Patton, P. M. Frank, R. N. Clark, Issues of Fault Diagnosis for Dynamic Systems, Springer London, London, 2000.

[3] I. Samy, I. Postlethwaite, D.-W. Gu, Survey and application of sensor fault detection and isolation schemes, Control Engineering Practice 19 (7) (2011) 658–674. doi:10.1016/j.conengprac.2011.03.002.

[4] M. Lal, R. Tiwari, Quantification of multiple fault parameters in flexible turbo-generator systems with incomplete rundown vibration data, Mechanical Systems and Signal Processing 41 (1-2) (2013) 546–563. doi:10.1016/j.ymssp.2013.06.025.

[5] Z.-S. Hou, Z. Wang, From model-based control to data-driven control: Survey, classification and perspective, Information Sciences 235 (2013) 3–35. doi:10.1016/j.ins.2012.07.014.

[6] P. Profos, T. Pfeiffer, Handbuch der industriellen Meßtechnik, 6. Auflage, Oldenbourg, München, 1994.

[7] ThyssenKrupp Steel Europe AG, Essernerstr. 244, 44793 Bochum (unveröffentlicht).

[8] Z. Feng, M. Liang, F. Chu, Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples, Mechanical Systems and Signal Processing 38 (1) (2013) 165–205. doi:10.1016/j.ymssp.2013.01.017.

[9] A. K. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing 20 (7) (2006) 1483–1510. doi:10.1016/j.ymssp.2005.09.012.

[10] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, D. Siegel, Prognostics and health

management design for rotary machinery systems—Reviews, methodology and applications, Mechanical Systems and Signal Processing 42 (2014) 314–334. doi:10.1016/j.ymssp.2013.06.004.

[11] J. Ma, J. Jiang, Applications of fault detection and diagnosis methods in nuclear power plants: A review, Progress in Nuclear Energy 53 (3) (2011) 255–266. doi:10.1016/j.pnucene.2010.12.001.

[12] M. Humberstone, B. Wood, J. Henkel, J. Hines, Differentiating between expanded and fault conditions using principal component analysis, Journal of Intelligent Manufacturing 23 (2) (2012) 179–188. doi:10.1007/s10845-009-0343-1.

[13] Y. Peng, M. Dong, M. J. Zuo, Current status of machine prognostics in condition-based maintenance: a review, The International Journal of Advanced Manufacturing Technology 50 (1-4) (2010) 297–313. doi:10.1007/s00170-009-2482-0.

[14] Bwag, Wien - Seestadt, SW-Areal 2013 (2).
URL https://commons.wikimedia.org/wiki/File:Wien_-_Seestadt,_SW-Areal_2013_(2).JPG,State:(01.Jul.2015).

[15] Kwerdenker, Kölnbreinspeicher.
URL https://commons.wikimedia.org/wiki/File:K%C3%B6lnbreinspeicher.jpg?uselang=de,State:(01.Jul.2015)

[16] M. Duhanic, 16199 dbtower duhanic.
URL https://commons.wikimedia.org/wiki/File:16199_dbtower_duhanic.jpg#/media/File:16199_dbtower_duhanic.jpg,State:(01.Jul.2015).

[17] H. Ortner, Mariazellerbahn 04.
URL https://commons.wikimedia.org/wiki/File:Mariazellerbahn_04.jpg,State:(01.Jul.2015).

[18] D. L. AG, Lufthansa Airbus A319-100 im Steigflug.
URL http://presse.lufthansa.com,(01.Jul.2015).

[19] A. Rother, M. Jelali, D. Söffker, A brief review and a first application of time-frequency-based analysis methods for monitoring of strip rolling mills, Journal of Process Control 35 (2015) 65–79.

[20] N. Julcher, Methoden der Fehlerdiagnose: eine Übersicht, Thesis, Eidgenössische Technische Hochschule Zürich, 2006.

[21] Z. Peng, F. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mechanical Systems and Signal Processing 18 (2) (2004) 199–221. doi:10.1016/S0888-3270(03)00075-X.

[22] W. Caesarendra, B. Kosasih, A. K. Tieu, C. A. S. Moodie, Circular domain features based condition monitoring for low speed slewing bearing, Mechanical Systems and Signal Processing 45 (1) (2013) 114–138. doi:10.1016/j.ymssp.2013.10.021.

[23] F. Serdio, E. Lughofer, K. Pichler, T. Buchegger, M. Pichler, H. Efendic, Fault detection in multi-sensor networks based on multivariate time-series models and orthogonal transformations, Information Fusion 20 (1) (2014) 272–291. doi:10.1016/j.inffus.2014.03.006.

[24] A. K. Nandi, C. Liu, M. L. D. Wong, Intelligent Vibration Signal Processing for Condition Monitoring, in: Proceedings of the International Conference Surveillance 7, Chartres, France, 2013, pp. 1–15.

[25] G. Box, G. Jenkins, G. Reinsel, Time Series Analysis: Forecasting and Control, Wiley, New Jersey, 1994.

[26] E. Brigham, The Fast Fourier Transform and its Applications, Prentice Hall, New Jersey, 1988.

[27] B. G. Ferguson, Application of the short-time Fourier transform and the Wigner–Ville distribution to the acoustic localization of aircraft, The Journal of the Acoustical Society of America 96 (2) (1994) 821. doi:10.1121/1.410320.

[28] W. Martin, P. Flandrin, Wigner-Ville spectral analysis of nonstationary processes, IEEE Transactions on Acoustics, Speech, and Signal Processing 33 (6) (1985) 1461–1470. doi:10.1109/TASSP.1985.1164760.

[29] S. G. Mallat, A Theory for Multiresolution Signal Decomposition : The Wavelet Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence II (7) (1989) 674–693.

[30] O. Rioul, M. Vetterli, Wavelets and signal processing, IEEE Signal Processing Magazine 8 (4) (1991) 14–38. doi:10.1109/79.91217.

[31] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, H. H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society

A: Mathematical, Physical and Engineering Sciences 454 (1971) (1998) 903–995. doi:10.1098/rspa.1998.0193.

[32] N. Huang, S. Shen, Hilbert–Huang Transform and its Applications, World Scientific, Singapore, 2005.

[33] I. Jolliffe, Principal Component Analysis, Springer Series in Statistics, Springer-Verlag, New York, 2002. doi:10.1007/b98835.

[34] Q. Guo, W. Wu, D. Massart, C. Boucon, S. de Jong, Feature selection in principal component analysis of analytical data, Chemometrics and Intelligent Laboratory Systems 61 (1-2) (2002) 123–132. doi:10.1016/S0169-7439(01)00203-9.

[35] D. Reynolds, A Gaussian mixture modeling approach to text-independent speaker identification, Ph.D. thesis, Georgia Institute of Technology (1992).

[36] D. W. Hosmer, S. Lemeshow, R. X. Sturdivant, Applied logistic regression, John Wiley & Sons, Inc., New York, 2000.

[37] J. Yan, J. Lee, Degradation Assessment and Fault Modes Classification Using Logistic Regression, Journal of Manufacturing Science and Engineering 127 (4) (2005) 912. doi:10.1115/1.1962019.

[38] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic press, San Diego, 1990.

[39] H. Sohn, C. R. Farrar, N. F. Hunter, K. Worden, Structural Health Monitoring Using Statistical Pattern Recognition Techniques, Journal of Dynamic Systems, Measurement, and Control 123 (4) (2001) 706. doi:10.1115/1.1410933.

[40] M. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/Non-Gaussian Bayesian tracking, IEEE Transactions on Signal Processing 50 (2) (2002) 174–188. doi:10.1109/78.978374.

[41] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering 82 (1) (1960) 35. doi:10.1115/1.3662552.

[42] S.J. Julier, J. Uhlmann, A new extension of the Kalman filter to nonlinear systems, in: Proceedings of the AeroSense: 11th International Symposium Aerospace/Defense Sensing, Simulation and Controls, 1997, pp. p. 182–193.

[43] T. Kohonen, The self-organizing map, Neurocomputing 21 (1-3) (1998) 1–6. doi:10.1016/S0925-2312(98)00030-7.

[44] T. Voegtlin, Recursive self-organizing maps, Neural Networks 15 (8-9) (2002) 979–991. doi:10.1016/S0893-6080(02)00072-2.

[45] F. Jensen, An Introduction to Bayesian Networks, Springer, New York, 1996.

[46] K. Murphy, Dynamic Bayesian networks: representation, inference and learning, Ph.D. thesis, University of California, Berkley (2002).

[47] P. Wasserman, Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York, 1989.

[48] J. C. D. Mandic, Recurrent Neural Networks for Prediction: Architectures, Learning Algorithms and Stability, Wiley, New York, 2001.

[49] G. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall, New Jersey, 1995.

[50] T. Ross, Fuzzy Logic with Engineering Applications, John Wiley & Sons, New York, 2004.

[51] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. doi:10.1007/BF00994018.

[52] C. J. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (2) (1998) 121–167. doi:10.1023/A:1009715923555.

[53] L. Rabiner, B. Juang, An introduction to hidden Markov models, IEEE ASSP Magazine 3 (1) (1986) 4–16. doi:10.1109/MASSP.1986.1165342.

[54] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (2) (1989) 257–286. doi:10.1109/5.18626.

[55] P. McFadden, M. Toozhy, Application of synchronous averaging to vibration monitoring of rolling element bearings, Mechanical Systems and Signal Processing 14 (2000) 891–906.

[56] C. Mechefske, J. Mathew, Fault detection and diagnosis in low speed rolling element bearings Part I: The use of parametric spectra, Mechanical Systems and Signal

Processing 6 (4) (1992) 297–307. doi:10.1016/0888-3270(92)90032-E.

[57] F. Choy, V. Polyshchuk, J. Zakrajsek, R. Handschuh, D. Townsend, Analysis of the effects of surface pitting and wear on the vibration of a gear transmission system, Tribology International 29 (1) (1996) 77–83. doi:10.1016/0301-679X(95)00037-5.

[58] L. S. Andrees, J. C. Rivadeneira, K. Gjika, C. Groves, G. LaRue, A Virtual Tool for Prediction of Turbocharger Nonlinear Dynamic Response: Validation Against Test Data, Journal of Engineering for Gas Turbines and Power 129 (4) (2007) 1035–1046. doi:10.1115/1.2436573.

[59] T.-W. Ha, Y.-B. Lee, C.-H. Kim, Leakage and rotordynamic analysis of a high pressure floating ring seal in the turbo pump unit of a liquid rocket engine, Tribology International 35 (3) (2002) 153–161. doi:10.1016/S0301-679X(01)00110-4.

[60] J. Sottile, F. Trutt, A. Leedy, Condition Monitoring of Brushless Three-Phase Synchronous Generators With Stator Winding or Rotor Circuit Deterioration, IEEE Transactions on Industry Applications 42 (5) (2006) 1209–1215. doi:10.1109/TIA.2006.880831.

[61] R. B. Randall, Applications of spectral kurtosis in machine diagnostics and prognostics, Key Engineering Materials 293-294 (2005) 21–32.

[62] C. Kar, Gearbox Health Monitoring through Multiresolution Fourier Transform of Vibration and Current Signals, Structural Health Monitoring, Technical Note 5 (2) (2006) 195–200. doi:10.1177/1475921706058002.

[63] W. Bartelmus, R. Zimroz, Vibration condition monitoring of planetary gearbox under varying external load, Mechanical Systems and Signal Processing 23 (1) (2009) 246–257. doi:10.1016/j.ymssp.2008.03.016.

[64] A. McCormick, A. Nandi, Neural network autoregressive modeling of vibrations for condition monitoring of rotating shafts, in: Proceedings of International Conference on Neural Networks (ICNN'97), Vol. 4, IEEE, pp. 2214–2218. doi:10.1109/ICNN.1997.614289.

[65] T. Sahraoui, S. Guessasma, N. Fenineche, G. Montavon, C. Coddet, Friction and wear behaviour prediction of HVOF coatings and electroplated hard chromium using neural computation, Materials Letters 58 (5) (2004) 654–660. doi:10.1016/j.matlet.2003.06.010.

[66] S. Vijayakumar, S. Muthukumar, Artificial neural network prediction and quantification of damage in impeller shaft using finite element simulation, International Journal of COMADEM 9 (1) (2006) 23–29.

[67] M. Jarrah, A. Al-Ali, Web-based monitoring and fault diagnostics of machinery, in: Proceedings of the IEEE International Conference on Mechatronics, 2004., IEEE, pp. 525–530. doi:10.1109/ICMECH.2004.1364494.

[68] O. A. Omitaomu, M. K. Jeong, A. B. Badiru, J. W. Hines, On-Line Prediction of Motor Shaft Misalignment Using Fast Fourier Transform Generated Spectra Data and Support Vector Regression, Journal of Manufacturing Science and Engineering 128 (4) (2006) 1019–1024. doi:10.1115/1.2194059.

[69] H.-W. Cho, Multivariate calibration for machine health monitoring: kernel partial least squares combined with variable selection, The International Journal of Advanced Manufacturing Technology 48 (5-8) (2010) 691–699. doi:10.1007/s00170-009-2309-z.

[70] Y. He, F. L. Chu, D. Guo, Detection and Configuration of the Shaft Crack in a Rotor-Bearing System by Genetic Algorithms, Key Engineering Materials 204-205 (2001) 221–230. doi:10.4028/www.scientific.net/KEM.204-205.221.

[71] B. Kim, S. Lee, M. Lee, J. Ni, J. Song, C. Lee, A comparative study on damage detection in speed-up and coast-down process of grinding spindle-typed rotor-bearing system, Journal of Materials Processing Technology 187-188 (2007) 30–36. doi:10.1016/j.jmatprotec.2006.11.222.

[72] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, Journal of Sound and Vibration 289 (4-5) (2006) 1066–1090. doi:10.1016/j.jsv.2005.03.007.

[73] J. H. Suh, S. R. Kumara, S. P. Mysore, Machinery Fault Diagnosis and Prognosis: Application of Advanced Signal Processing Techniques, CIRP Annals - Manufacturing Technology 48 (1) (1999) 317–320. doi:10.1016/S0007-8506(07)63192-8.

[74] L. Gao, Research on Fault Diagnosis Technology of Low Speed and Heavy Duty Equipments Based on Wavelet Analysis, Chinese Journal of Mechanical Engineering 41 (12) (2005) 222. doi:10.3901/JME.2005.12.222.

[75] H.-R. Li, B.-H. Xu, Fault prognosis of hydraulic pump in the missile launcher, Acta Armamentarii 30 (7) (2009) 900–906.

[76] F. Wan, Q. Xu, S. Li, Vibration analysis of cracked rotor sliding bearing system with rotor–stator rubbing by harmonic wavelet transform, Journal of Sound and Vibration 271 (3-5) (2004) 507–518. doi:10.1016/S0022-460X(03)00277-3.

[77] Z. Wang, H. Jiang, Robust incipient fault diagnosis methods for enhanced aircraft engine rotor prognostics, in: Proceedings of the Second International Conference on Innovative Computing, Information and Control, 2007, pp. 455–458.

[78] W. Zanardelli, E. Strangas, S. Aviyente, Failure prognosis for permanent magnet AC drives based on wavelet analysis, in: Proceedings of the IEEE International Conference on Electric Machines and Drives, 2005, pp. 64–70.

[79] H. Xie, G.Wen, Long-term vibration trend prediction of rotor system state based on support vector regression and DiscreteWavelet Decomposition, in: Proceedings of the 2009 International Workshop on Intelligent Systems and Applications, 2009, pp. 1–4.

[80] V. Rai, A. Mohanty, Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform, Mechanical Systems and Signal Processing 21 (6) (2007) 2607–2615. doi:10.1016/j.ymssp.2006.12.004.

[81] B. Liu, S. Riemenschneider, Y. Xu, Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum, Mechanical Systems and Signal Processing 20 (3) (2006) 718–734. doi:10.1016/j.ymssp.2005.02.003.

[82] H. Li, Y. Zhang, H. Zheng, Wear detection in gear system using Hilbert-Huang transform, Journal of Mechanical Science and Technology 20 (11) (2006) 1781–1789. doi:10.1007/BF03027572.

[83] D. Brie, M. Tomczak, H. Oehlmann, A. Richard, Gear Crack Detection by Adaptive Amplitude and Phase Demodulation, Mechanical Systems and Signal Processing 11 (1) (1997) 149–167. doi:10.1006/mssp.1996.0068.

[84] X. Zhang, C. Xu, S. Liang, Q. Xie, L. Haynes, An integrated approach to bearing fault diagnostics and prognostics, in: Proceedings of the American Control Conference (AAC), 2005, pp. 2750–2755.

[85] Y. Chen, L. Lan, A fault detection technique for air-source heat pump water chiller/heaters, Energy and Buildings 41 (8) (2009) 881–887. doi:10.1016/j.enbuild.2009.03.007.

[86] M. E. Orchard, G. J. Vachtsevanos, A particle-filtering approach for on-line fault diagnosis and failure prognosis, Transactions of the Institute of Measurement and Control 31 (3-4) (2009) 221–246. doi:10.1177/0142331208092026.

[87] J.-D. Wu, C.-W. Huang, R. Huang, An application of a recursive Kalman filtering algorithm in rotating machinery fault diagnosis, NDT & E International 37 (5) (2004) 411–419. doi:10.1016/j.ndteint.2003.11.006.

[88] Y. Zhan, V. Makis, A. K. Jardine, Adaptive state detection of gearboxes under varying load conditions based on parametric modelling, Mechanical Systems and Signal Processing 20 (1) (2006) 188–221. doi:10.1016/j.ymssp.2004.08.004.

[89] R. Houser, J. Sorenson, J. Harianto, H. Wijaya, M. Satyanarayana, Comparison of analytical predictions with dynamic noise and vibration measurements for a simple idler gearbox, VDI Berichte 2 (1665) (2002) 995–1002.

[90] S. Yang, An experiment of state estimation for predictive maintenance using Kalman filter on a DC motor, Reliability Engineering & System Safety 75 (1) (2002) 103–111. doi:10.1016/S0951-8320(01)00107-7.

[91] R. Huang, L. Xi, X. Li, C. Richard Liu, H. Qiu, J. Lee, Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods, Mechanical Systems and Signal Processing 21 (1) (2007) 193–207. doi:10.1016/j.ymssp.2005.11.008.

[92] R. C. M. Yam, P. Tse, L. Li, P. Tu, Intelligent Predictive Decision Support System for Condition-Based Maintenance, The International Journal of Advanced Manufacturing Technology 17 (5) (2001) 383–391. doi:10.1007/s001700170173.

[93] P. Wang, G. Vachtsevanos, Fault prognostics using dynamic wavelet neural networks, AI EDAM 15 (04) (2001) 349–365.

[94] M. R. Dellomo, Helicopter Gearbox Fault Detection: A Neural Network Based Approach, Journal of Vibration and Acoustics 121 (3) (1999) 265. doi:10.1115/1.2893975.

[95] W. J. Staszewski, K. Worden, Classification of faults in gearboxes ? pre-processing algorithms and neural networks, Neural Computing & Applications 5 (3) (1997) 160–183. doi:10.1007/BF01413861.

[96] C. Byington, M. Watson, D. Edwards, Data-driven neural network methodology

to remaining life predictions for aircraft actuator components, in: 2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720), Vol. 6, IEEE, pp. 3581–3589. doi:10.1109/AERO.2004.1368175.

[97] T. Khawaja, G. Vachtsevanos, B. Wu, Reasoning about uncertainty in prognosis: a confidence prediction neural network approach, in: Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society, 2005, pp. 7–12.

[98] E. Liang, R. J. Rodriguez, A. A. Husseiny, Prognostics/diagnostics of mechanical equipment by neural network, Neural Networks 1 (1) (1988) 33–41.

[99] M. Gibiec, Prediction of Machines Health with Application of an Intelligent Approach – a Mining Machinery Case Study, Key Engineering Materials 293-294 (2005) 661–668. doi:10.4028/www.scientific.net/KEM.293-294.661.

[100] T. Engin, Prediction of relative efficiency reduction of centrifugal slurry pumps: empirical- and artificial-neural network-based methods, Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy 221 (1) (2007) 41–50. doi:10.1243/09576509JPE224.

[101] J. Penman, Feasibility of using unsupervised learning, artificial neural networks for the condition monitoring of electrical machines, IEE Proceedings - Electric Power Applications 141 (6) (1994) 317. doi:10.1049/ip-epa:19941263.

[102] F. Filippetti, G. Franceschini, C. Tassoni, Neural networks aided on-line diagnostics of induction motor rotor faults, IEEE Transactions on Industry Applications 31 (4) (1995) 892–899. doi:10.1109/28.395301.

[103] C. Byington, M. Watson, D. Edwards, Dynamic signal analysis and neural network modeling for life prediction of flight control actuators, in: 60th Annual Forum Proceedings - American Helicopter Society, 2004, pp. 928–937.

[104] S. Pandit, S. Wu, Time Series and System Analysis with Applications, John Wiley & Sons, 1983.

[105] F. Galati, B. Forrester, S. Dey, Application of the generalised likelihood ratio algorithm to the detection of a bearing fault in a helicopter transmission, in: Australian Journal of Mechanical Engineering, 2008, pp. 169–175.

[106] G. Wang, Z. Luo, X. Qin, Y. Leng, T. Wang, Fault identification and classification of

rolling element bearing based on time-varying autoregressive spectrum, Mechanical Systems and Signal Processing 22 (4) (2008) 934–947. doi:10.1016/j.ymssp.2007.10.008.

[107] W. Wang, A. K. Wong, Autoregressive Model-Based Gear Fault Diagnosis, Journal of Vibration and Acoustics 124 (2) (2002) 172–179. doi:10.1115/1.1456905.

[108] Z. S. Chen, Y. M. Yang, Z. Hu, G. J. Shen, Detecting and Predicting Early Faults of Complex Rotating Machinery Based on Cyclostationary Time Series Model, Journal of Vibration and Acoustics 128 (5) (2006) 666. doi:10.1115/1.2345674.

[109] X. Wang, V. Makis, Autoregressive model-based gear shaft fault diagnosis using the Kolmogorov–Smirnov test, Journal of Sound and Vibration 327 (3) (2009) 413–423. doi:10.1016/j.jsv.2009.07.004.

[110] B. Sinha, Trend prediction from steam turbine responses of vibration and eccentricity, Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy 216 (1) (2002) 97–104.

[111] B. Satish, N. Sarma, A Fuzzy BP approach for diagnosis and prognosis of bearing faults in induction motors, in: Proceedings of the IEEE Power Engineering Society General Meeting, 2005, pp. 2291–2294.

[112] P. J. Dempsey, A. A. Afjeh, Integrating Oil Debris and Vibration Gear Damage Detection Technologies Using Fuzzy Logic, Journal of the American Helicopter Society 49 (2) (2004) 109. doi:10.4050/JAHS.49.109.

[113] A. Sözen, E. Arcaklioğlu, A. Erisen, M. Akçayol, Performance prediction of a vapour-compression heat-pump, Applied Energy 79 (3) (2004) 327–344. doi:10.1016/j.apenergy.2003.12.013.

[114] S. Perovic, P. Unsworth, E. Higham, Fuzzy logic system to detect pump faults from motor current spectra, in: Conference Record of the 2001 IEEE Industry Applications Conference. 36th IAS Annual Meeting (Cat. No.01CH37248), Vol. 1, IEEE, pp. 274–280. doi:10.1109/IAS.2001.955423.

[115] R. Sepe, J. Miller, A.R. Gale, Intelligent efficiency mapping of a hybrid electric vehicle starter/alternator using fuzzy logic, in: Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, 1999, pp. 8–12.

[116] P. Vas, AI-based Electrical Machines and Drives: Application of Fuzzy, Neural,

Fuzzy-Neural, and Genetic-Algorithm-Based Techniques, Oxford University Press, New York, 1999.

[117] F. Filippetti, G. Franceschini, C. Tassoni, P. Vas, Recent developments of induction motor drives fault diagnosis using AI techniques, IEEE Transactions on Industrial Electronics 47 (5) (2000) 994–1004. doi:10.1109/41.873207.

[118] J. Liu, D. Djurdjanovic, J. Ni, N. Casoetto, J. Lee, Similarity based method for manufacturing process performance prediction and diagnosis, Computers in Industry 58 (6) (2007) 558–566. doi:10.1016/j.compind.2006.12.004.

[119] J. Yang, Y. Zhang, Y. Zhu, Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension, Mechanical Systems and Signal Processing 21 (5) (2007) 2012–2024. doi:10.1016/j.ymssp.2006.10.005.

[120] B. Samata, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mechanical Systems and Signal Processing2004 18 625–644.

[121] H. Ocak, K. A. Loparo, F. M. Discenzo, Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics, Journal of Sound and Vibration 302 (4-5) (2007) 951–961. doi:10.1016/j.jsv.2007.01.001.

[122] X. Wu, Y. Li, T. Lundell, A. Guru, Integrated prognosis of AC servo motor driven linear actuator using hidden semi-Markov models, in: Proceedings of the IEEE International Electric Machines and Drives Conference, 2009, pp. 1408–1413.

[123] W. Wang, A model to predict the residual life of rolling element bearings given monitored condition information to date, IMA Journal of Management Mathematics 13 (1) (2002) 3–16. doi:10.1093/imaman/13.1.3.

[124] Y. LI, T. KURFESS, S. LIANG, Stochastic Prognostics for Rolling Element Bearings, Mechanical Systems and Signal Processing 14 (5) (2000) 747–762. doi:10.1006/mssp.2000.1301.

[125] F. Z. Feng, D. D. Zhu, P. C. Jiang, H. Jiang, GA-SVR Based Bearing Condition Degradation Prediction, Key Engineering Materials 413-414 (2009) 431–437. doi:10.4028/www.scientific.net/KEM.413-414.431.

[126] Z. Li, Z. He, Y. Zi, H. Jiang, Rotating machinery fault diagnosis using signal-

adapted lifting scheme, Mechanical Systems and Signal Processing 22 (3) (2008) 542–556. doi:10.1016/j.ymssp.2007.09.008.

[127] Y. Lei, Z. He, Y. Zi, Q. Hu, Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs, Mechanical Systems and Signal Processing 21 (5) (2007) 2280–2294. doi:10.1016/j.ymssp.2006.11.003.

[128] S. Loutridis, Damage detection in gear systems using empirical mode decomposition, Engineering Structures 26 (12) (2004) 1833–1841. doi:10.1016/j.engstruct.2004.07.007.

[129] J.-Z.Wang, G.-H. Zhou, X.-S. Zhao, L. S.-X., Gearbox fault diagnosis and prediction based on empirical mode decomposition scheme,, in: Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, 2007, pp. 1072–1075.

[130] W. Yang, P. Tavner, Empirical mode decomposition, an adaptive approach for interpreting shaft vibratory signals of large rotating machinery, Journal of Sound and Vibration 321 (3-5) (2009) 1144–1170. doi:10.1016/j.jsv.2008.10.012.

[131] F. Wu, L. Qu, Diagnosis of subharmonic faults of large rotating machinery based on EMD, Mechanical Systems and Signal Processing 23 (2) (2009) 467–475. doi:10.1016/j.ymssp.2008.03.007.

[132] D. Stringer, P. Sheth, P. Allaire, Gear modeling methodologies for advancing prognostic capabilities in rotary-wing transmission systems, in: American Helicopter Society 64th Annual Forum - AHS, 2008, pp. 1492–1504.

[133] C. Stoisser, S. Audebert, A comprehensive theoretical, numerical and experimental approach for crack detection in power plant rotating machinery, Mechanical Systems and Signal Processing 22 (4) (2008) 818–844. doi:10.1016/j.ymssp.2007.11.013.

[134] B.-S. Yang, S. Kwon Jeong, Y.-M. Oh, A. C. C. Tan, Case-based reasoning system with Petri nets for induction motor fault diagnosis, Expert Systems with Applications 27 (2) (2004) 301–311. doi:10.1016/j.eswa.2004.02.004.

[135] P. Chen, M. Taniguchi, T. Toyota, Z. He, Fault diagnosis method for machinery in unsteady operating condition by instantaneous power spectrum and genetic programming, Mechanical Systems and Signal Processing 19 (1) (2005) 175–194. doi:10.1016/j.ymssp.2003.11.004.

[136] D.-M. Yang, A. Stronach, P. MacConnell, J. Penman, Third-Order Spectral Techniques for the Diagnosis of Motor Bearing Condition Using Artificial Neural Networks, Mechanical Systems and Signal Processing 16 (2-3) 391–411. doi:10.1006/mssp.2001.1469.

[137] L. Wang, P. Ye, J. Wang, S. Yang, Bispectrum characteristics of the faults of rubbing rotor system based on experimental study, Journal of Vibration Engineering 15 (2002) 339–334.

[138] W. Q. Wang, M. Golnaraghi, F. Ismail, Prognosis of machine health condition using neuro-fuzzy systems, Mechanical Systems and Signal Processing 18 (4) (2004) 813–831. doi:10.1016/S0888-3270(03)00079-7.

[139] H. Esen, M. Inalli, A. Sengur, M. Esen, Modelling a ground-coupled heat pump system using adaptive neuro-fuzzy inference systems, International Journal of Refrigeration 31 (1) (2008) 65–74. doi:10.1016/j.ijrefrig.2007.06.007.

[140] K. R. Al-Balushi, B. Samanta, Gear fault diagnosis using energy-based features of acoustic emission signals, Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 216 (3) (2002) 249–263. doi:10.1177/095965180221600304.

[141] W. Jiang, S. K. Spurgeon, J. A. Twiddle, F. S. Schlindwein, Y. Feng, S. Thanagasundram, A wavelet cluster-based band-pass filtering and envelope demodulation approach with application to fault diagnosis in a dry vacuum pump, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 221 (11) (2007) 1279–1286. doi:10.1243/09544062JMES544.

[142] M. Benbouzid, M. Vieira, C. Theys, Induction motors' faults detection and localization using stator current advanced signal processing techniques, IEEE Transactions on Power Electronics 14 (1) (1999) 14–22. doi:10.1109/63.737588.

[143] X.-G. Hou, Z.-G. Wu, L. Xia, Method for detecting rotor faults in asynchronous motors based on the square of the Park's vector modulus, in: Proceedings of the Chinese Society of Electrical Engineering, Vol. 23, 2003, pp. 137–140.

[144] R. Schoen, B. Lin, T. Habetler, J. Schlag, S. Farag, An unsupervised, on-line system for induction motor fault detection using stator current monitoring, IEEE Transactions on Industry Applications 31 (6) (1995) 1280–1286. doi:10.1109/28.475698.

[145] R. J. Bankert, V. K. Singh, H. Rajiyah, Model based diagnostics and prognosis sys-

tem for rotating machinery, in: American Society of Mechanical Engineers, Houston, USA, 1995, pp. 5–9.

[146] N. Arthur, J. Penman, Induction machine condition monitoring with higher order spectra, IEEE Transactions on Industrial Electronics 47 (5) (2000) 1031–1041. doi:10.1109/41.873211.

[147] H. Nejjari, M. Benbouzid, Monitoring and diagnosis of induction motors electrical faults using a current Park's vector pattern learning approach, IEEE Transactions on Industry Applications 36 (3) (2000) 730–735. doi:10.1109/28.845047.

[148] J. Pittner, A Useful Control Model for Tandem Hot Metal Strip Rolling, IEEE Transactions on Industry Applications 46 (6) (2010) 2251–2258.

[149] J. Pittner, Controller for improving the quality of the tandem rolling of hot metal strip, in: American Control Conference (AAC), 2010, pp. 6095–6100.

[150] S. K. Yildiz, J. F. Forbes, B. Huang, Y. Zhang, F. Wang, V. Vaculik, M. Dudzic, Dynamic modelling and simulation of a hot strip finishing mill, Applied Mathematical Modelling 33 (7) (2009) 3208–3225. doi:10.1016/j.apm.2008.10.035.

[151] X.-H. Jiao, L.-P. Shao, Y. Peng, Adaptive Coordinated Control for Hot Strip Finishing Mills, International Journal of Iron and Steel Research 18 (4) (2011) 36–43. doi:10.1016/S1006-706X(11)60047-2.

[152] W.-g. Li, Z.-h. Guo, J. Yi, X.-h. Liu, Optimization of Roll Shifting Strategy of Alternately Rolling in Hot Strip Mill, Journal of Iron and Steel Research, International 19 (5) (2012) 37–42. doi:10.1016/S1006-706X(12)60097-1.

[153] A. Steinboeck, D. Wild, A. Kugi, Nonlinear model predictive control of a continuous slab reheating furnace, Control Engineering Practice 21 (4) (2013) 495–508. doi:10.1016/j.conengprac.2012.11.012.

[154] S. J. Mannanal, S. Dehnhardt, A. Gesser, T. Eickmeyer, Global adaptive model for prediction, characterisation and damping of vibration in hot strip mills, Research Fund Coal and Steel, Luxembourg , 2011.

[155] A. A. Kuldiwar, Finite Element Modeling of Strip Curvature During Hot Rolling, in: Proceedings of 9th International LS-DYNA Users Conference, no. 2, 2006, pp. 17–23.

[156] E. Brusa, L. Lemma, D. Benasciutti, Vibration analysis of a Sendzimir cold rolling mill and bearing fault detection, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 224 (8) (2010) 1645–1654. doi:10.1243/09544062JMES1540.

[157] F. J. García, I. Díaz, I. Álvarez, D. Pérez, D. G. Ordonez, M. Domínguez, Time-Frequency Analysis of hot rolling using maifold lerning, in: Engineering Applications of Neural Networks, 2011, pp. 150–155.

[158] Y. Wang, Z. He, J. Xiang, Y. Zi, Application of local mean decomposition to the surveillance and diagnostics of low-speed helical gearbox, Mechanism and Machine Theory 47 (2012) 62–73. doi:10.1016/j.mechmachtheory.2011.08.007.

[159] K. Peng, K. Zhang, G. Li, D. Zhou, Contribution rate plot for nonlinear quality-related fault diagnosis with application to the hot strip mill process, Control Engineering Practice 21 (4) (2013) 360–369. doi:10.1016/j.conengprac.2012.11.013.

[160] L. Hui, T. Chaonan, P. Kaixiang, Data-driven Modeling and Online Algorithm for Hot Rolling Process, in: Proceedings of the 30th Chinese Control Conference, Yantai, China, 2011, pp. 1560–1564.

[161] G.-Y. Li, M. Dong, A Wavelet and Neural Networks Based on Fault Diagnosis for HAGC System of Strip Rolling Mill, International Journal of Iron and Steel Research 18 (1) (2011) 31–35. doi:10.1016/S1006-706X(11)60007-1.

[162] S. Lesecq, S. Gentil, S. Taleb, Fault detection based on wavelet transform. Application to a roughing mill, in: Proceedings of IFAC Fault Detection, Supervision and Safety of Technical Processes, Beijing, 2006, pp. 1115–1120.

[163] F. Serdio, E. Lughofer, K. Pichler, T. Buchegger, H. Efendic, Residual-based fault detection using soft computing techniques for condition monitoring at rolling mills, Information Sciences 259 (2014) 304–320. doi:10.1016/j.ins.2013.06.045.

[164] J. Li, X. Chen, Z. He, Adaptive stochastic resonance method for impact signal detection based on sliding window, Mechanical Systems and Signal Processing 36 (2) (2013) 240–255. doi:10.1016/j.ymssp.2012.12.004.

[165] J. Li, X. Chen, Z. He, Multi-stable stochastic resonance and its application research on mechanical fault diagnosis, Journal of Sound and Vibration 332 (22) (2013) 5999–6015. doi:10.1016/j.jsv.2013.06.017.

[166] J. Yuan, Z. He, Y. Zi, H. Liu, Gearbox fault diagnosis of rolling mills using multiwavelt sliding window neighboring coefficient denoising and optimal blind deconvolution, Science in China Series E: Technological Sciences 52 (10) (2009) 2801–2809. doi:10.1007/s11431-009-0253-7.

[167] Y. Chen, Y. Zi, H. Cao, Z. He, H. Sun, A data-driven threshold for wavelet sliding window denoising in mechanical fault detection, Science China: Technological Sciences 57 (3) (2014) 589–597. doi:10.1007/s11431-013-5451-7.

[168] A. M. Lukasson-Herzig, Optimierung der Stahlbandgeometrie im Hinblick auf den Bandsäbel in Warmbreitbandstrassen, VDM Verlag, Saarbrücken, 2008.

[169] R. Nandan, R. Rai, R. Jayakanth, S. Moitra, N. Chakraborti, A. Mukhopadhyay, Regulating Crown and Flatness During Hot Rolling: A Multiobjective Optimization Study Using Genetic Algorithms, Materials and Manufacturing Processes 20 (3) (2005) 459–478. doi:10.1081/AMP-200053462.

[170] J. F. Liu, M. Chen, J. Y. Gu, L. Cheng, Remote Fault Diagnosis System Based on EMD and SVM for Heavy Rolling-Mills, Advanced Materials Research 889-890 (2014) 681–686. doi:10.4028/www.scientific.net/AMR.889-890.681.

[171] F. Sanfilippo, E. Musella, N. D. Biase, A. Petrelli, An advanced internet based data reporting system for on line monitoring of a hot rolling mill, in: Proceedings of the 7. Conferenza Associazione Italiana Metallurgia, 2001, pp. 1–8.

[172] Z.-M. Chen, F. Luo, Y.-G. Xu, W. Yu, Roll Eccentricity Compensation Based on Anti-Aliasing Wavelet Analysis Method, Journal of Iron and Steel Research, International 16 (2) (2009) 35–39. doi:10.1016/S1006-706X(09)60024-8.

[173] S. G. Mallat, S. Zhong, Characterization of Signals from Multiscale Edges, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (1992) 701–732.

[174] E. Arinton, S. Caraman, J. Korbicz, Neural networks for modelling and fault detection of the inter-stand strip tension of a cold tandem mill, Control Engineering Practice 20 (7) (2012) 684–694. doi:10.1016/j.conengprac.2012.03.007.

[175] A. Debón, J. Carlos Garcia-Díaz, Fault diagnosis and comparing risk for the steel coil manufacturing process using statistical models for binary data, Reliability Engineering & System Safety 100 (2012) 102–114. doi:10.1016/j.ress.2011.12.022.

[176] X. Zhang, X. Liu, Cascade Control for Hydraulic Automatic Gauge Control of

Hot Rolling Mills Based on Data-driven Theory, in: International Conference on Computer Science and Service Systems, Vol. 1, 2014, pp. 434–437.

[177] L. Wang, Y. Yuan, Y. Shao, A method of chatter marks identification based on Autocorrelation-threshold, Key Engineering Materials 572 (2014) 485–488.

[178] DIN 8583-2:2003-09, Manufacturing processes forming under compressive conditions - Part 2: Rolling; Classification, subdivision, terms and definitions, Beuth Verlag, 2003.

[179] A. Rother, M. Jelali, D. Söffker, Signal-based Fault Prognosis Approach Based on Time-Frequency Analysis Applied to Industrial Data, in: IWSHM 10th Interantional Workshop on Structural Health Monitoring, Stanford, USA, 2015.

[180] A. Rother, M. Jelali, D. Söffker, Entwicklung eines Verfahrens zur Fehlerdiagnose mittels Support Vector Machine auf Basis von gemessenen Betriebsdaten, in: 10. Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung (AKIDA), Aachen, Germany, 2014, pp. 33–40.

[181] A. Rother, M. Jelali, D. Söffker, Development of a Fault Detection Approach Based on SVM Applied to Industrial Data, in: EWSHM-7th European Workshop on Structural Health Monitoring, Nantes, Fance, 2014.

[182] K. Lange, M. Liewald, Umformtechnik Handbuch für Industrie und Wissenschaft - Band 2: Massivumformung, Springer, Heidelberg, 1988.

[183] R. Kopp, H. Wiegels, Einführung in die Umformtechnik, Mainz, Aachen, 1999.

[184] H. Palkowski, M. Albedyhl, G. Füsers, VdEh, Surface defects in hot rolled flat steel products, Verlag Stahl Eisen GmbH, Düsseldorf, 1996.

[185] A. Rother, Interaktionsanalyse von Anstellungs- und Walzkraftsignalen beim Warmbandwalzen, Masterarbeit, Hochschule Niederrhein, Krefeld, 2011.

[186] ABB, Millmate Roll Force Systems mit Millmate Controller 400 Benutzerhandbuch, 2007.

[187] R. Kohavi, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, International Joint Conference on Artificial Intelligence 14 (1995) 1137–1143. doi:10.1067/mod.2000.109031.

[188] L. Cohen, Time-Frequency Distributions-A, Proceedings of the IEEE 77 (7) (1989) 941–981.

[189] S. Schlagner, U. Strehlau, Fourier-Analyse versus Wavelet-Analyse, PAMM 5 (1) (2005) 125–126. doi:10.1002/pamm.200510043.

[190] A. Grossmann, J. Morlet, Decomposition of Hardy Function into Square Integrable Wavelets of Constant Shape, SIAM J. Math. Anal. 15 (4) (1984) 723–736.

[191] C. K. Chui, An Introduction to Wavelets, Academic Press Inc., San Diego, 1992.

[192] I. Daubechies, C. Heil, Ten Lectures on Wavelets, Vol. 6, Society for Industrial & Applied Mathematics, 1992.

[193] Y. Meyer, Wavelets and Operators, Cambridge University Press, 1992.

[194] J. Lin, L. Qu, Feature Extraction Based on Morlet Wavelet and Its Application for Mechanical Fault Diagnosis, Journal of Sound and Vibration 234 (1) (2000) 135–148. doi:10.1006/jsvi.2000.2864.

[195] H. Sun, Z. He, Y. Zi, J. Yuan, X. Wang, J. Chen, S. He, Multiwavelet transform and its applications in mechanical fault diagnosis – A review, Mechanical Systems and Signal Processing 43 (1-2) (2014) 1–24. doi:10.1016/j.ymssp.2013.09.015.

[196] R. Yan, R. X. Gao, X. Chen, Wavelets for fault diagnosis of rotary machines: A review with applications, Signal Processing 96 (2014) 1–15. doi:10.1016/j.sigpro.2013.04.015.

[197] F. Auger, P. Flandrin, Improving the readability of time-frequency and time-scale representations by the reassignment method, IEEE Transactions On Signal Processing 43 (5) (1995) 1068–1089.

[198] C. Smith, C. M. Akujuobi, P. Hamory, K. Kloesel, An approach to vibration analysis using wavelets in an application of aircraft health monitoring, Mechanical Systems and Signal Processing 21 (2007) 1255–1320.

[199] F. Al-Badour, M. Sunar, L. Cheded, Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques, Mechanical Systems and Signal Processing 25 (6) (2011) 2083–2101. doi:10.1016/j.ymssp.2011.01.017.

[200] J. O. Strömberg, A modied Franklin system and higher-order spline systems on Rn

as unconditional bases for Hardy spaces, in: Conf. on Harmonic Analysis in Honor of A. Zygmund, Vol. II, 1983, pp. 475–494.

[201] C. E. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal 27 (1948) (1948) 379–423, 623–656.

[202] A. Oppenheim, R. Schafer, J. Buck, Discrete-Time Signal Processing, 3rd Edition, Prentice Hall, New Jersey, 1999.

[203] Z. Yao, D. Mei, Z. Chen, On-line chatter detection and identification based on wavelet and support vector machine, Journal of Materials Processing Technology 210 (5) (2010) 713–719. doi:10.1016/j.jmatprotec.2009.11.007.

[204] D. Luczak, Mechanical resonance frequencies identyfication of direct drive using wavelet analysis, in: Proceedings of 17th Methods and Models in Automation and Robotics, 2012, pp. 29–32.

[205] G. Cai, X. Chen, Z. He, Sparsity-enabled signal decomposition using tunable Q-factor wavelet transform for fault feature extraction of gearbox, Mechanical Systems and Signal Processing 41 (1-2) (2013) 34–53. doi:10.1016/j.ymssp.2013.06.035.

[206] A. J. Joshi, F. Porikli, S. Member, N. Papanikolopoulos, Scalable Active Learning for Multi-Class Image Classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (11) (2012) 2259–2273. doi:10.1109/TPAMI.2012.21.

[207] D. Keren, M. Werman, J. Feinberg, A Probabilistic Approach to Pattern Matching in the Continuous Domain, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (10) (2012) 1873–1885.

[208] M. Lamraoui, M. Thomas, M. El Badaoui, Cyclostationarity approach for monitoring chatter and tool wear in high speed milling, Mechanical Systems and Signal Processing 44 (1-2) (2014) 177–198. doi:10.1016/j.ymssp.2013.05.001.

[209] V. Climente-Alarcona, J. Antonino-Daviua, M. Riera-Guaspa, R. Puche-Panaderoa, L. Escobar, Application of the Wigner-Ville distribution for the detection of rotor asymmetries and eccentricity through high-order harmonics, Electric Power Systems Research 91 (2012) 28–36.

[210] Y. Lei, J. Lin, Z. He, M. J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mechanical Systems and Signal Processing 35 (1-2) (2013) 108–126. doi:10.1016/j.ymssp.2012.09.015.

[211] C. Bisu, L. Olteanu, R. Laheurte, P. Darnis, O. Cahuc, Experimental Approach on Torsor Dynamic Analysis for Milling Process Monitoring and Diagnosis, Procedia CIRP 12 (2013) 73–78. doi:10.1016/j.procir.2013.09.014.

[212] G. Georgoulas, T. Loutas, C. D. Stylios, V. Kostopoulos, Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition, Mechanical Systems and Signal Processing 41 (1-2) (2013) 510–525. doi:10.1016/j.ymssp.2013.02.020.

[213] Z. Peng, P. W. Tse, F. Chu, A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing, Mechanical Systems and Signal Processing 19 (5) (2005) 974–988. doi:10.1016/j.ymssp.2004.01.006.

[214] S. Abe, Support Vector Machines for Pattern Classification, 2nd Edition, Springer Verlag London, London, 2010.

[215] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (3) (2011) 1–27.

[216] L. Fahrmeir, R. Künstler, I. Pigeot, G. Tutz, Statistik, Springer, Berlin Heidelberg New York, 2007.

[217] C. Nachtigall, M. Wirtz, Wahrscheinlichkeitsrechnung und Interferenzstatistik, Band 2, Beltz Juventa, Weinheim, München, 2013.

[218] Q. McNemar, Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages, Psychometrica 122 (1947) 153–157.

[219] A. L. Edwards, Note on the "correction for continuity" in testing the significance of the difference between correlated proportions, Psychometrica 13 (3) (1948) 185–187.

A Journal papers and conference contributions

This thesis is based on the results and development steps published in the following publications and/or presented at the corresponding conferences.

Journal article

[19] A. Rother, M. Jelali, D. Söffker, A brief review and a first application of time-frequency-based analysis methods for monitoring of strip rolling mills, Journal of Process Control 35 (2015) 65-79.

Conference papers

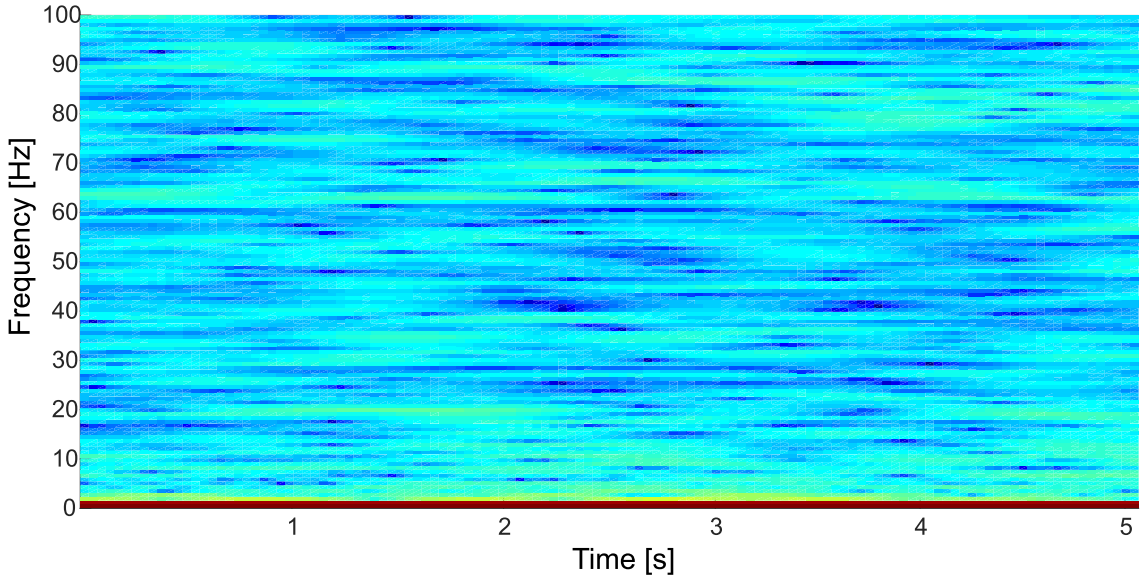
[179] A. Rother, M. Jelali, D. Söffker: Signal-based Fault Prognosis Approach Based on Time-Frequency Analysis Applied to Industrial Data, in: IWSHM 10th Interantional Workshop on Structural Health Monitoring, Stanford, USA, September 1-3, 2015.

[180] A. Rother, M. Jelali, D. Söffker: Entwicklung eines Verfahrens zur Fehlerdiagnose mittels Support Vector Machine auf Basis von gemessenen Betriebsdaten, in: 10. Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung AKIDA, Aachen, Germany, November 19-20, 2014, pp. 33–40.

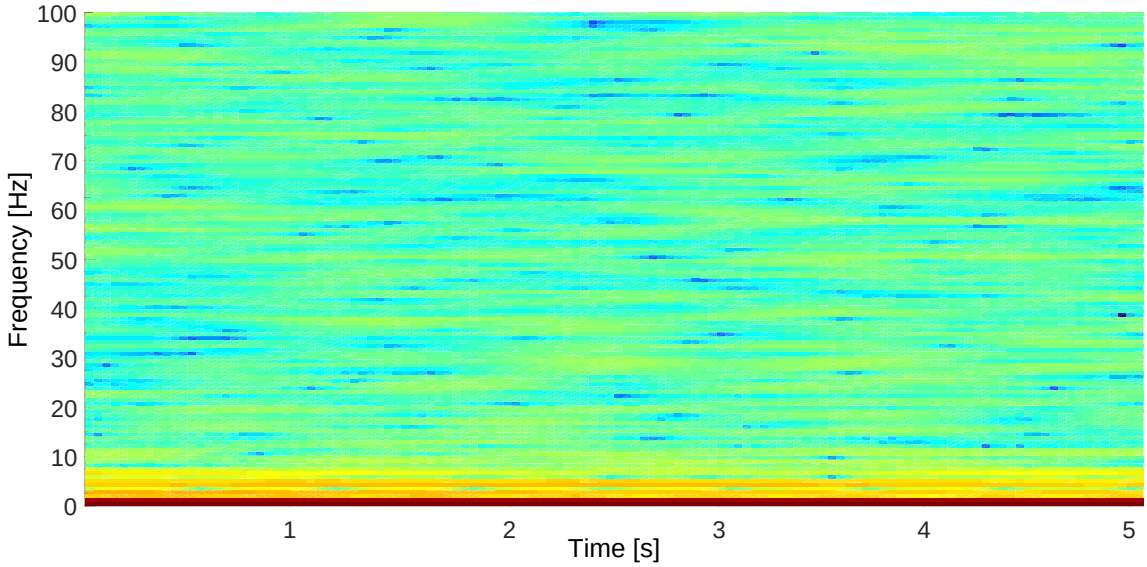
[181] A. Rother, M. Jelali, D. Söffker: Development of a Fault Detection Approach Based on SVM Applied to Industrial Data, in: EWSHM-7th European Workshop on Structural Health Monitoring, Nantes, France, July 8-11, 2014.

B Appendix

The graphical representation of the application results of the five methods applied for feature extraction are given in Figures B.1-B.10. Figures B.11-B.12 visualize the correlation coefficients of EMD-CC for State 1-4. Figures B.13-B.14 illustrate the results of EMD-CC applied for fault prognosis.

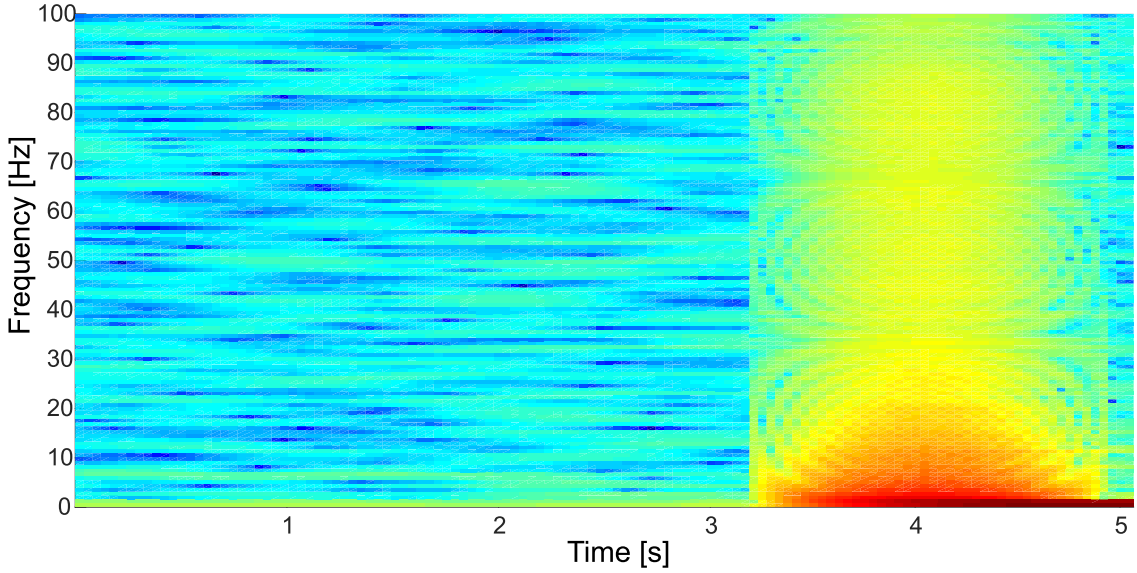


(a) Application of STFT to fault-free case State 1 [19]

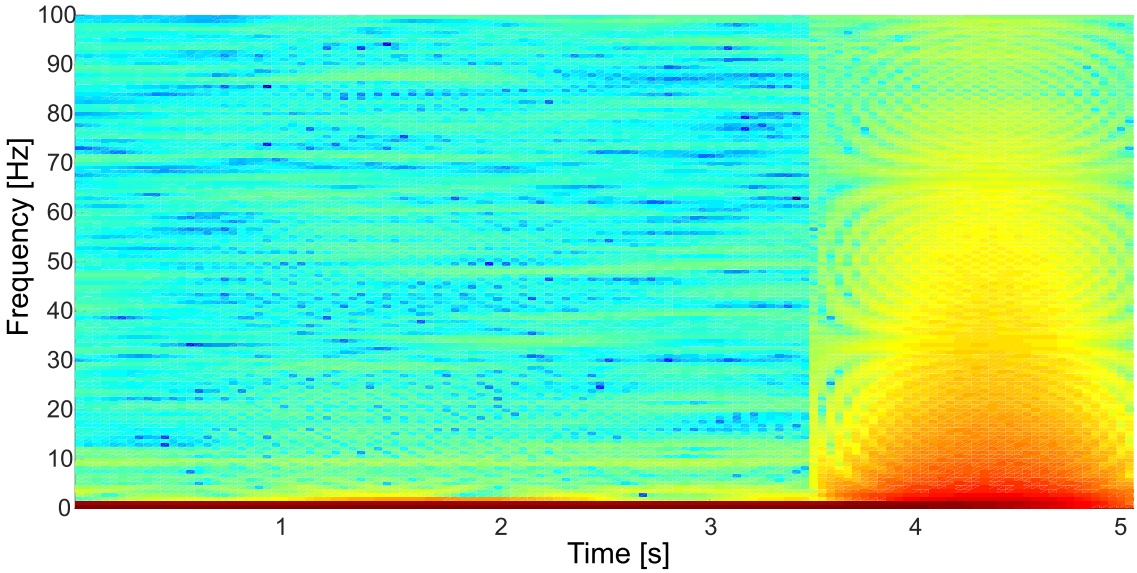


(b) Application of STFT to fault-free case State 2

Figure B.1: Graphical results of STFT

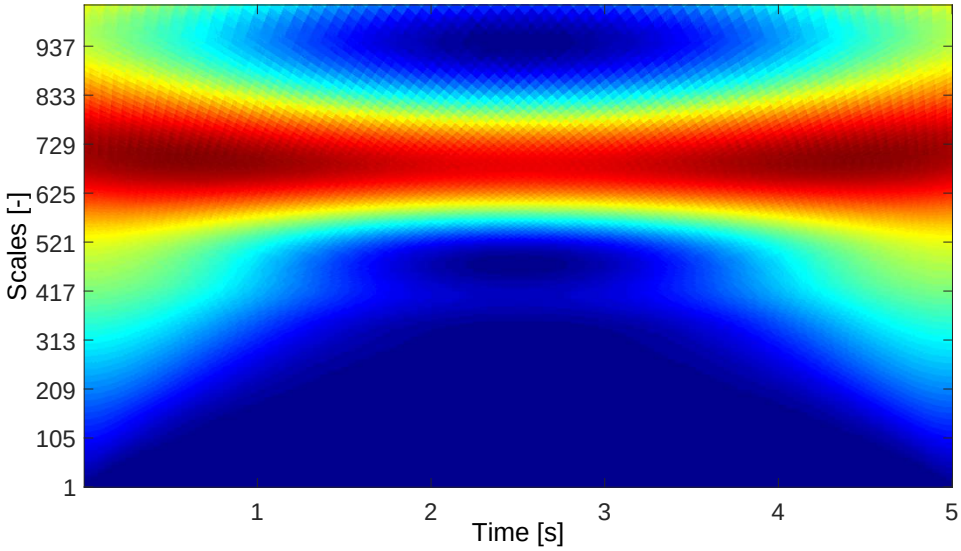


(a) Application of STFT to fault case State 3 [19]

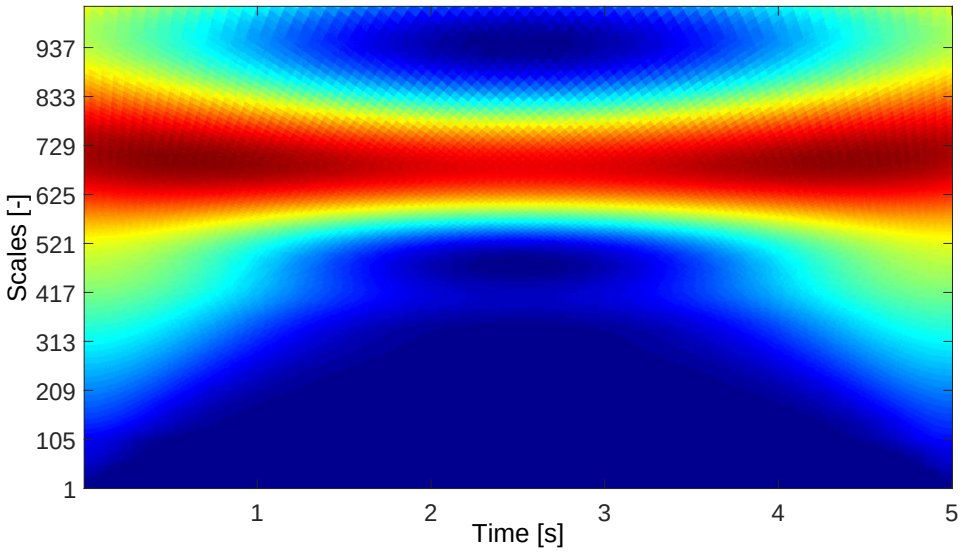


(b) Application of STFT to fault case State 4 [19]

Figure B.2: Graphical results of STFT

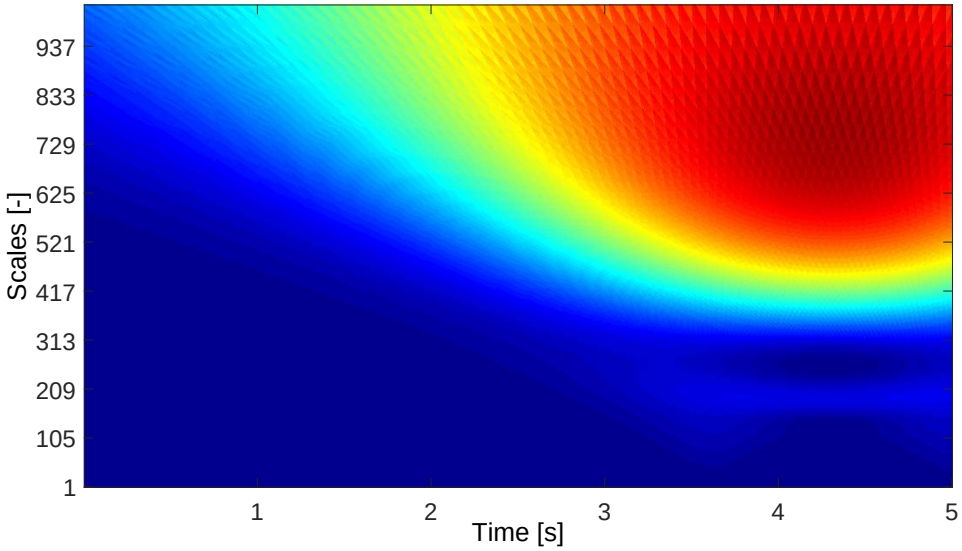


(a) Application of CWT to fault-free case State 1 [19]

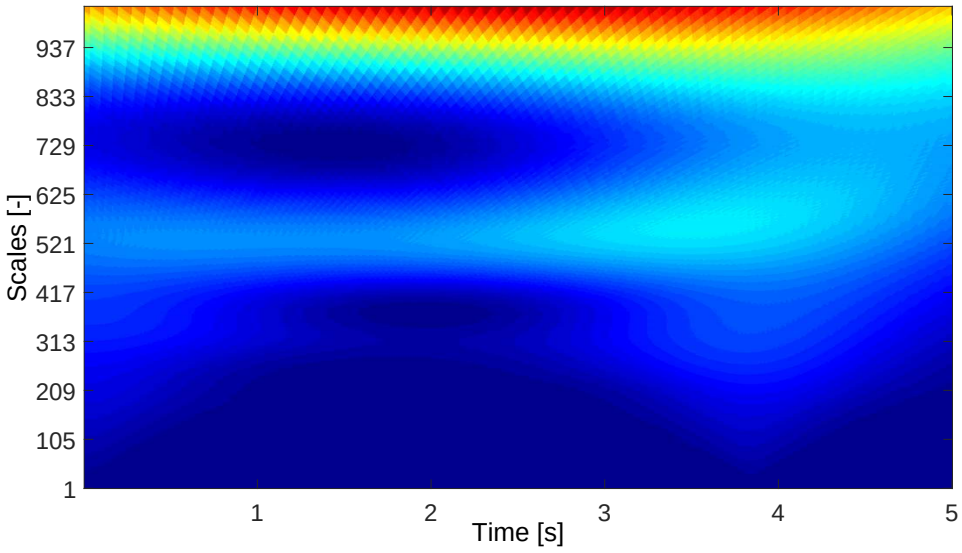


(b) Application of CWT to fault-free case State 2

Figure B.3: Graphical results of CWT

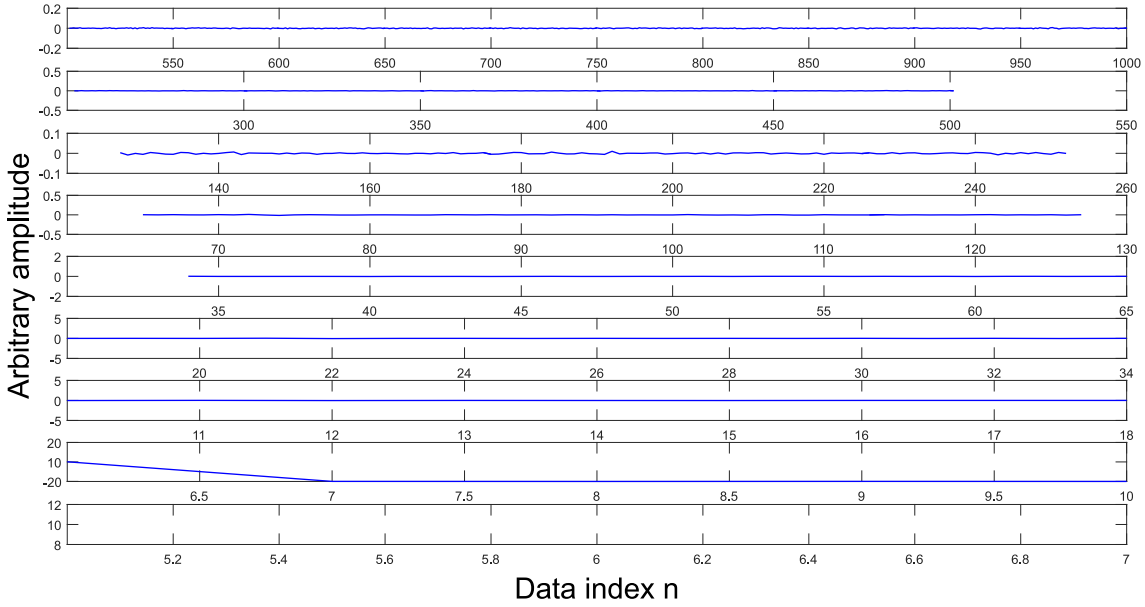


(a) Application of CWT to fault case State 3 [19]

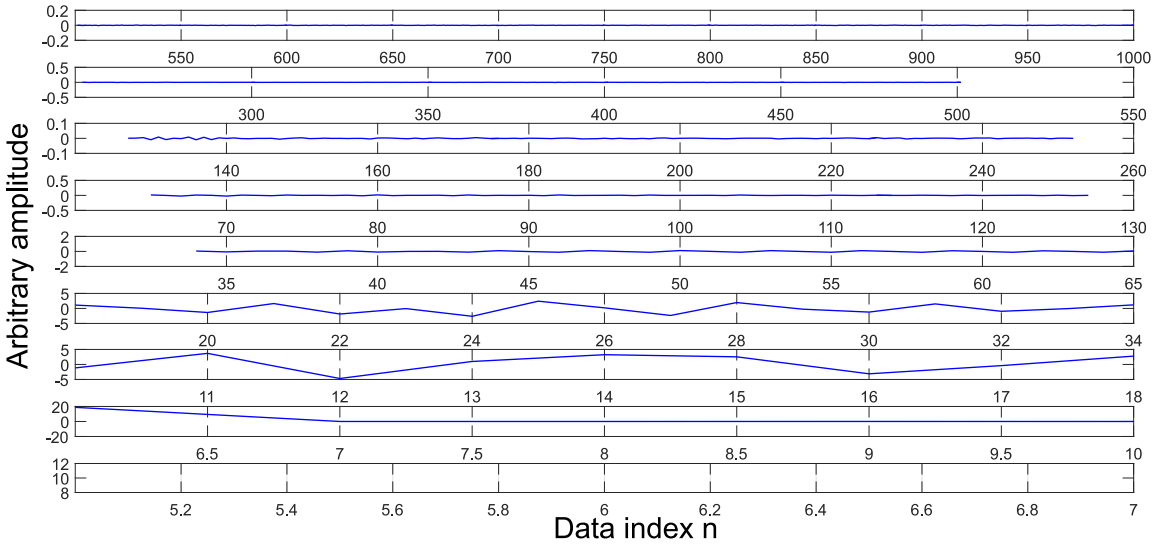


(b) Application of CWT to fault case State 4 [19]

Figure B.4: Graphical results of CWT

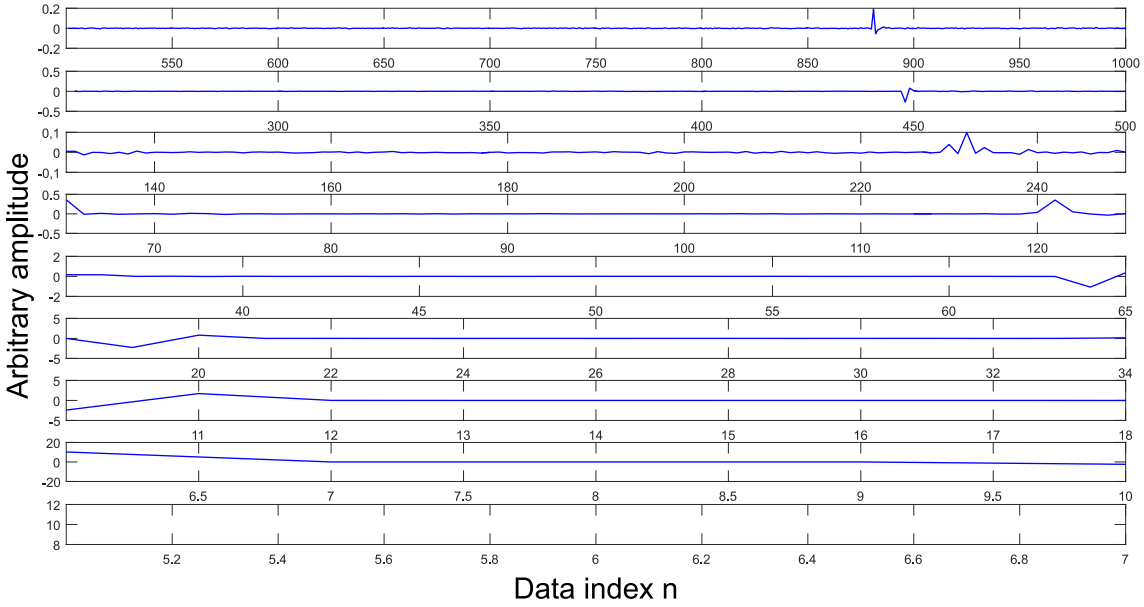


(a) Application of DWT to fault-free case State 1 [19]

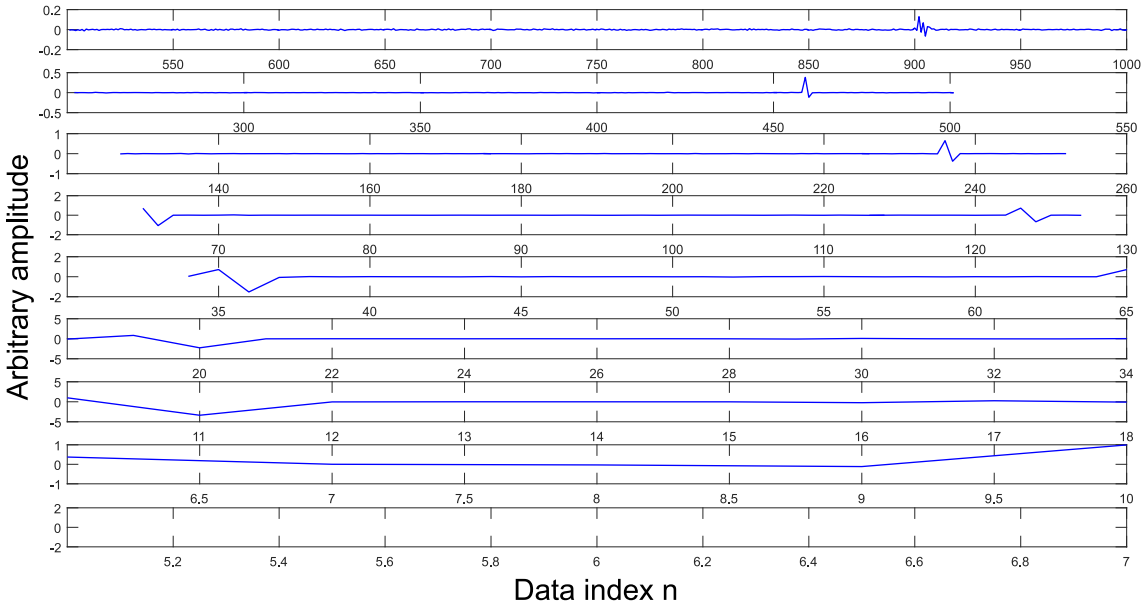


(b) Application of DWT to fault-free case State 2

Figure B.5: Graphical results of DWT

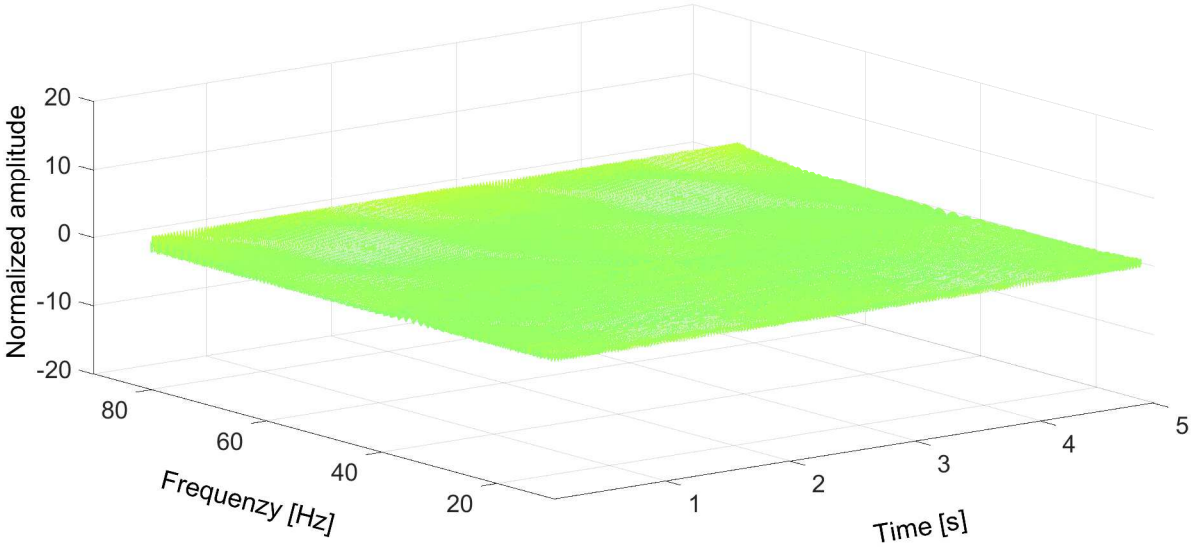


(a) Application of DWT to fault case State 3 [19]

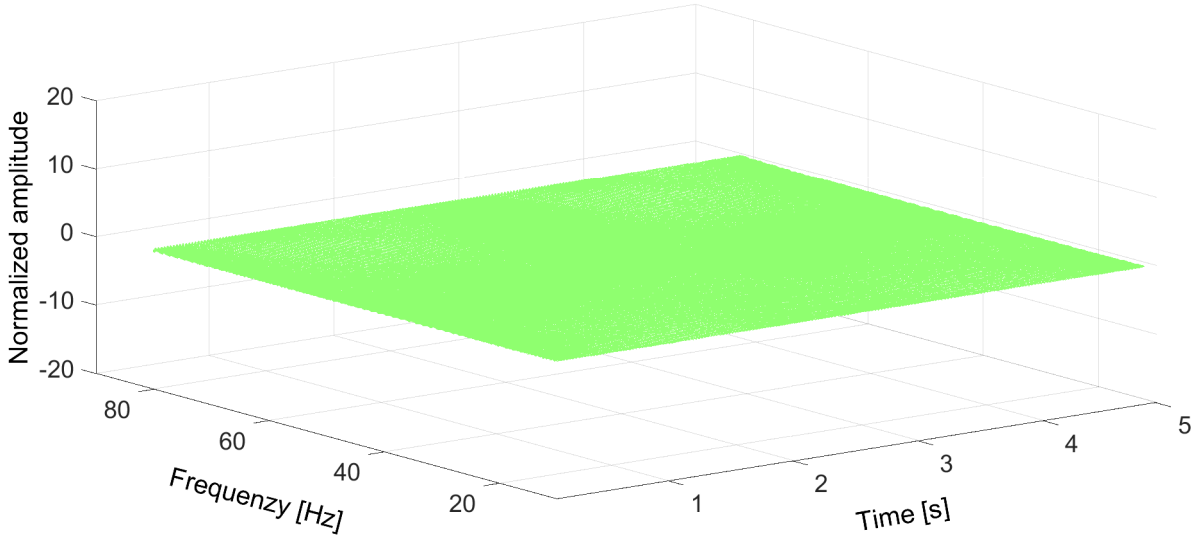


(b) Application of DWT to fault case State 4 [19]

Figure B.6: Graphical results of DWT

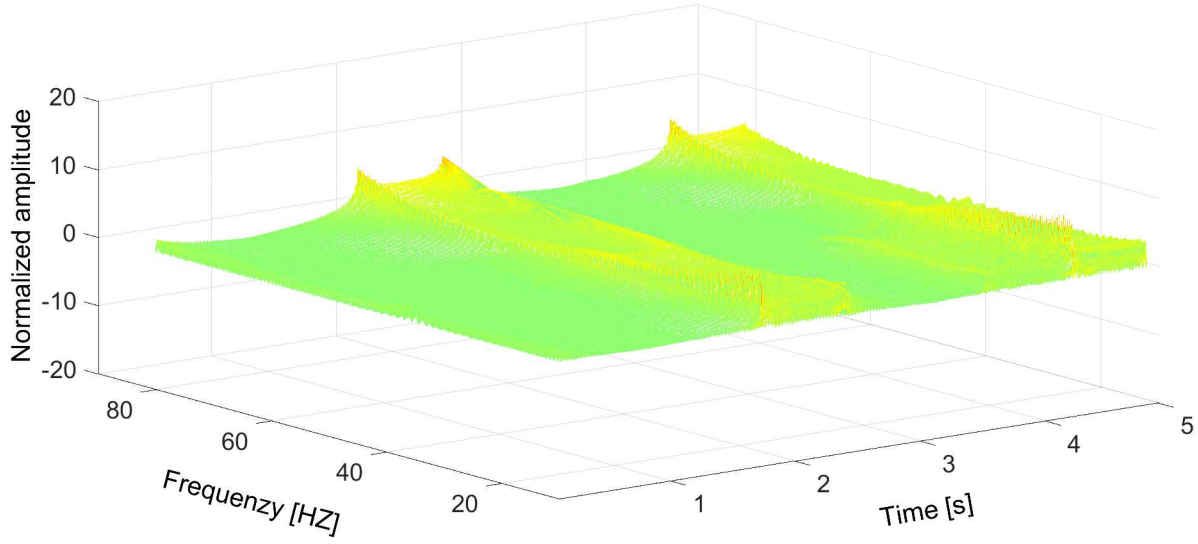


(a) Application of WVD to fault-free case State 1 [19]

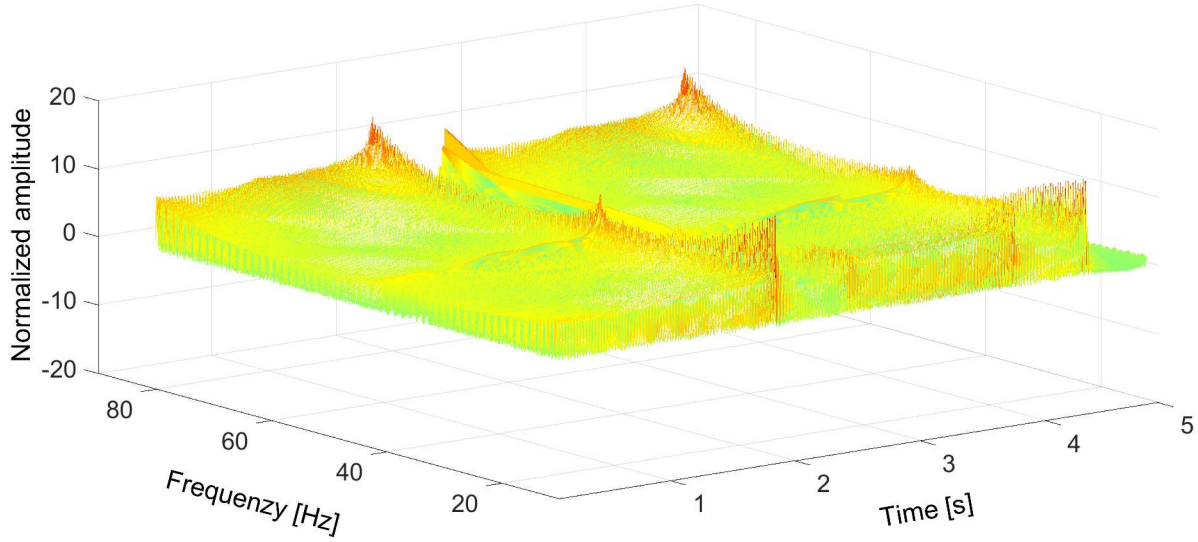


(b) Application of WVD to fault-free case State 2

Figure B.7: Graphical results of WVD

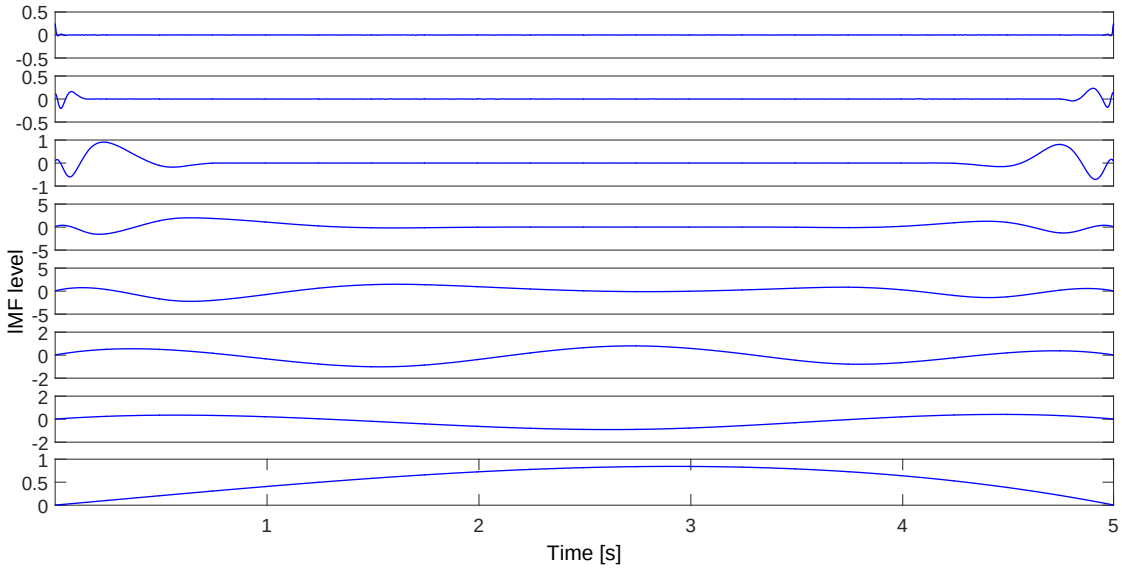


(a) Application of WVD to fault case State 3 [19]

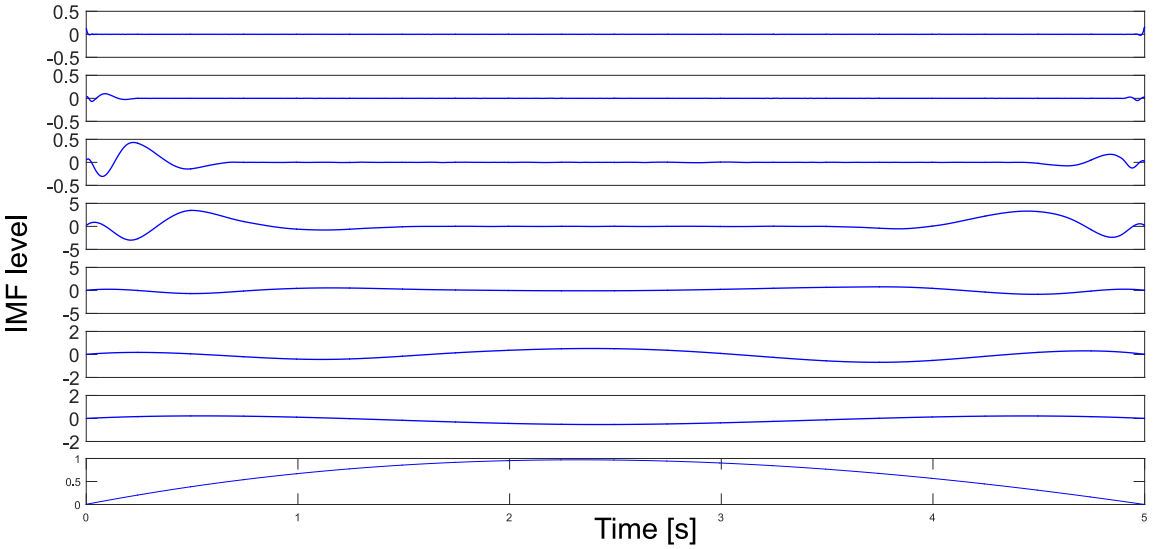


(b) Application of WVD to fault case State 4 [19]

Figure B.8: Graphical results of WVD

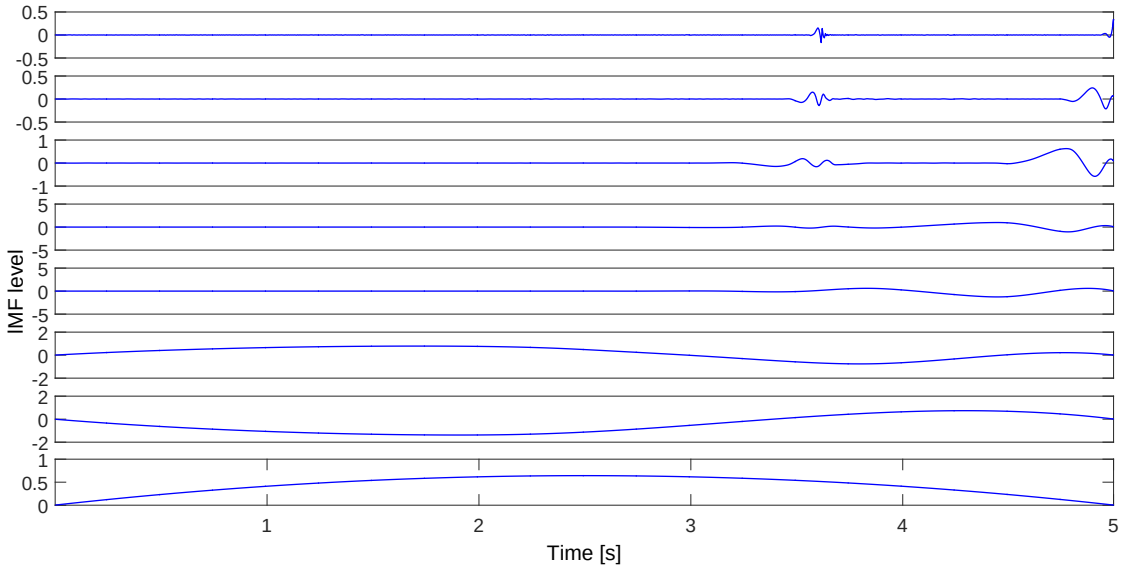


(a) Application of EMD to fault-free case State 1 [19]

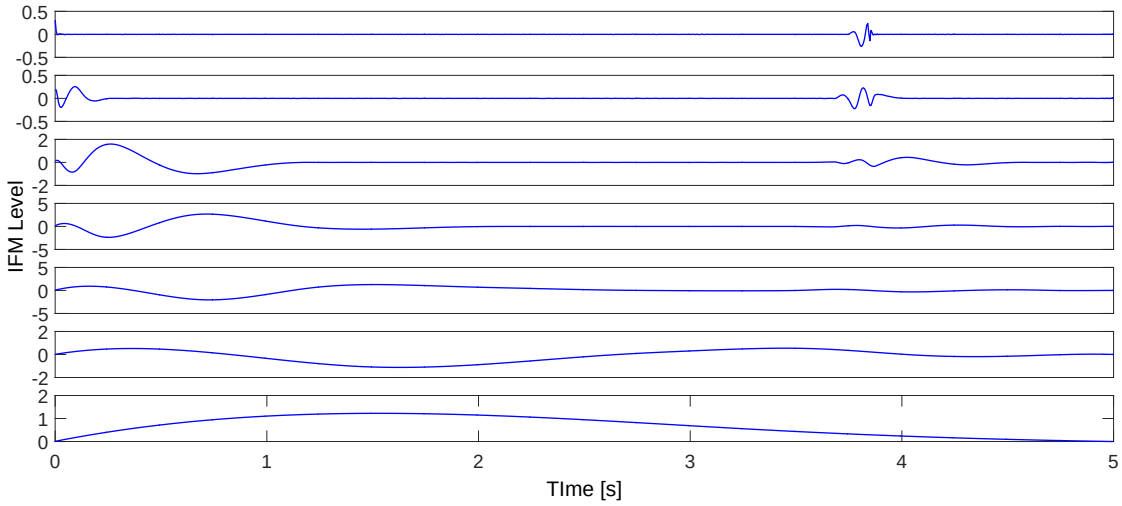


(b) Application of EMD to fault-free case State 2

Figure B.9: Graphical results of EMD

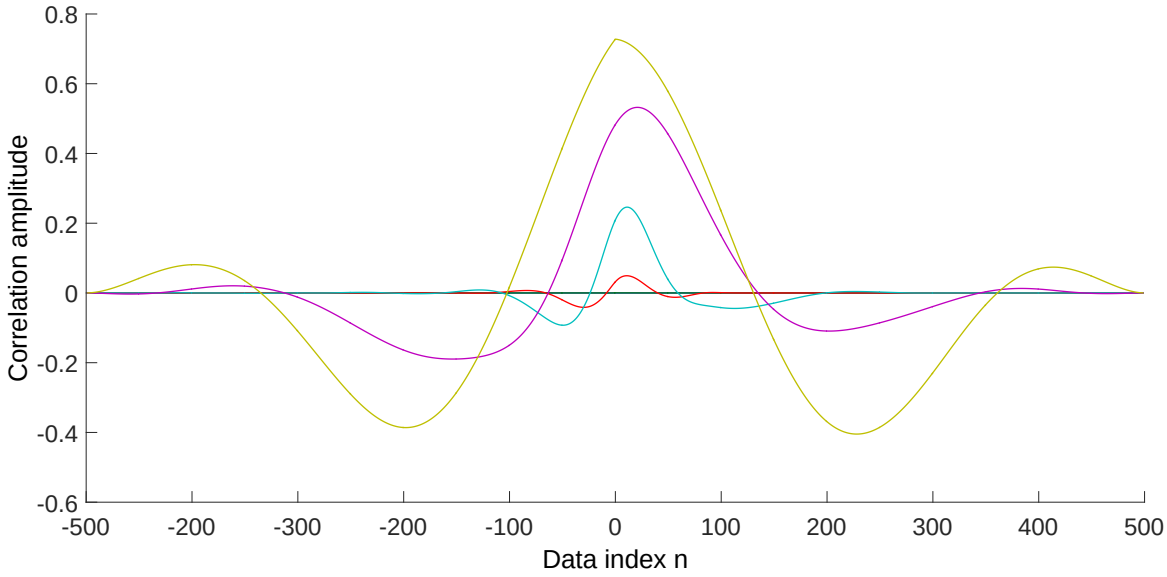


(a) Application of EMD to fault case State 3 [19]

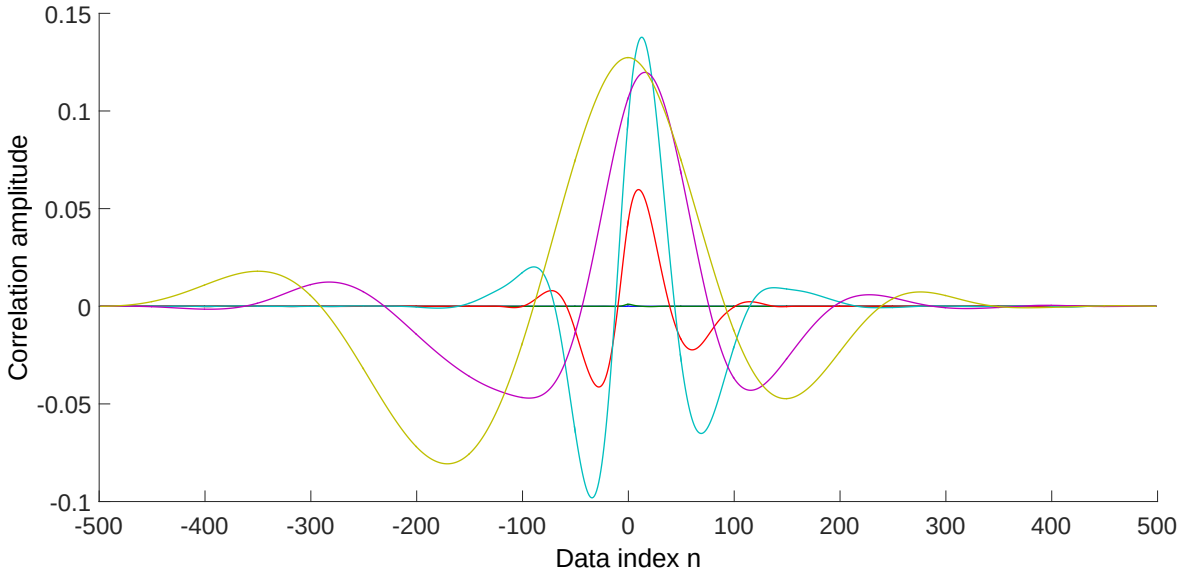


(b) Application of EMD to fault case State 4 [19]

Figure B.10: Graphical results of EMD

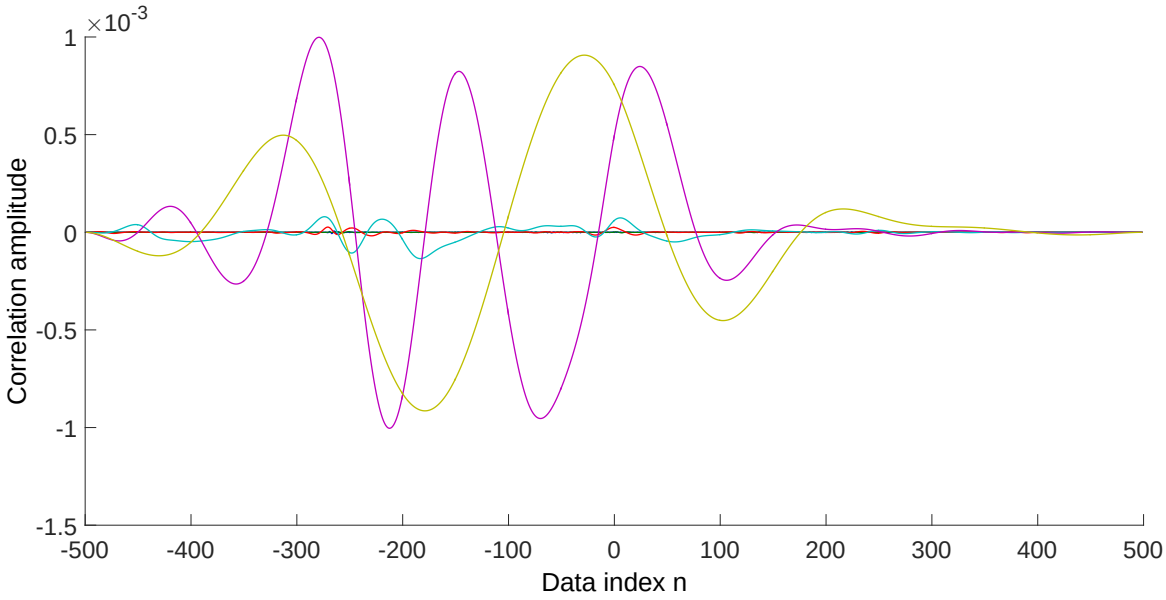


(a) Application of EMD-CC to fault-free case State 1

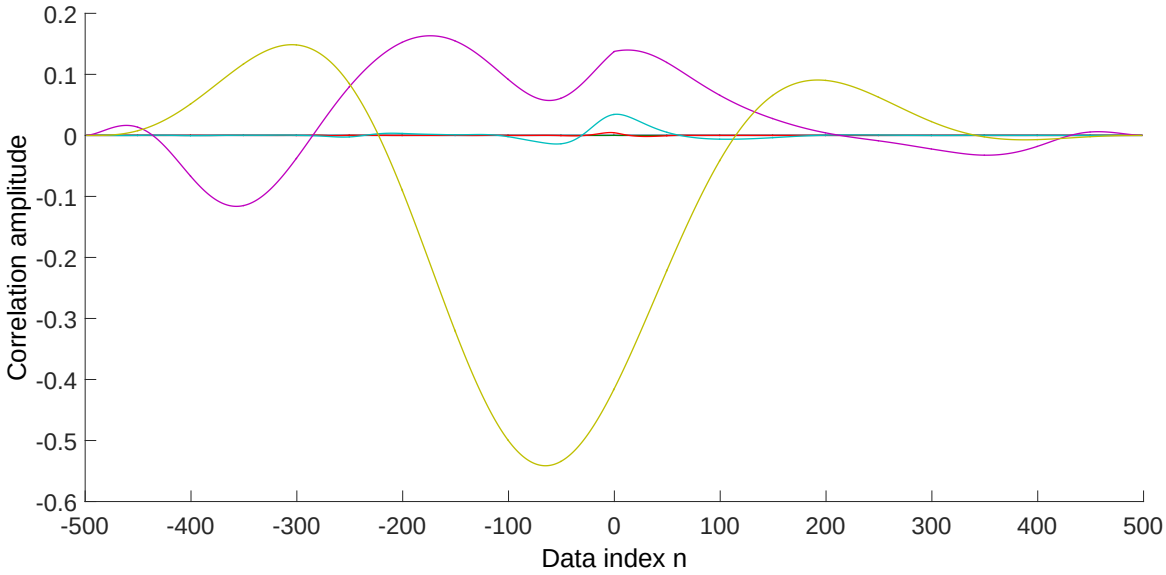


(b) Application of EMD-CC to fault-free case State 1

Figure B.11: Graphical results of EMD-CC

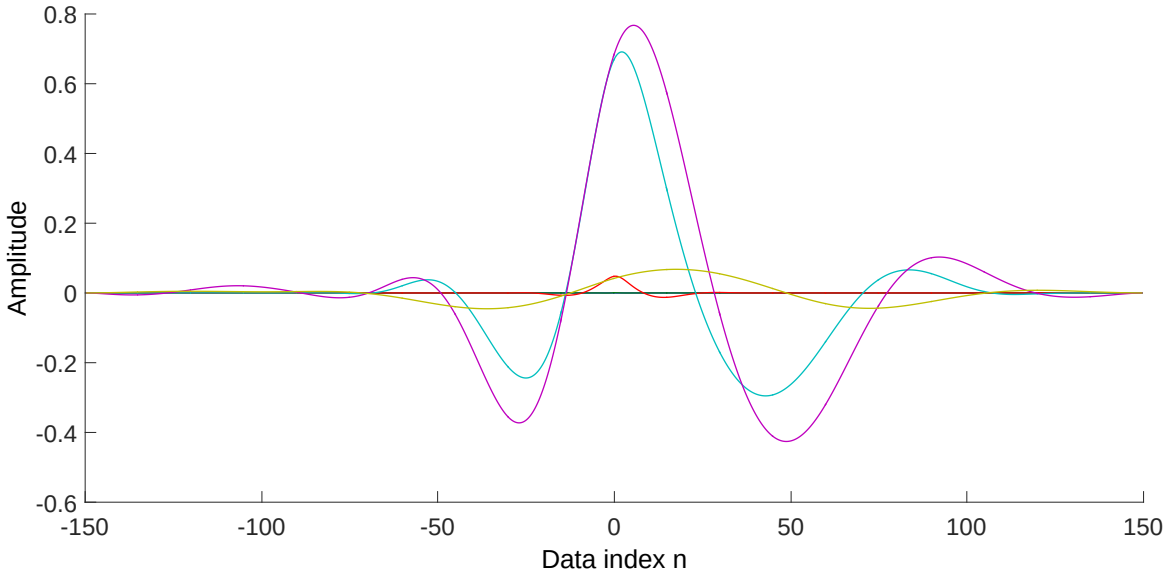


(a) Application of EMD-CC to fault case State 3

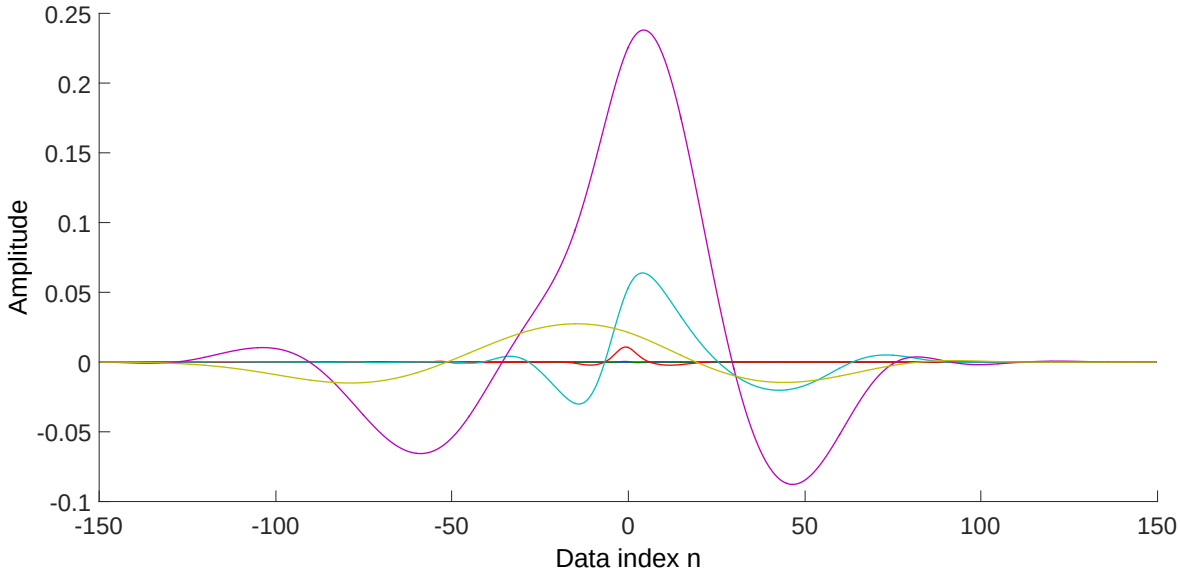


(b) Application of EMD-CC to fault case State 4

Figure B.12: Graphical results of EMD-CC

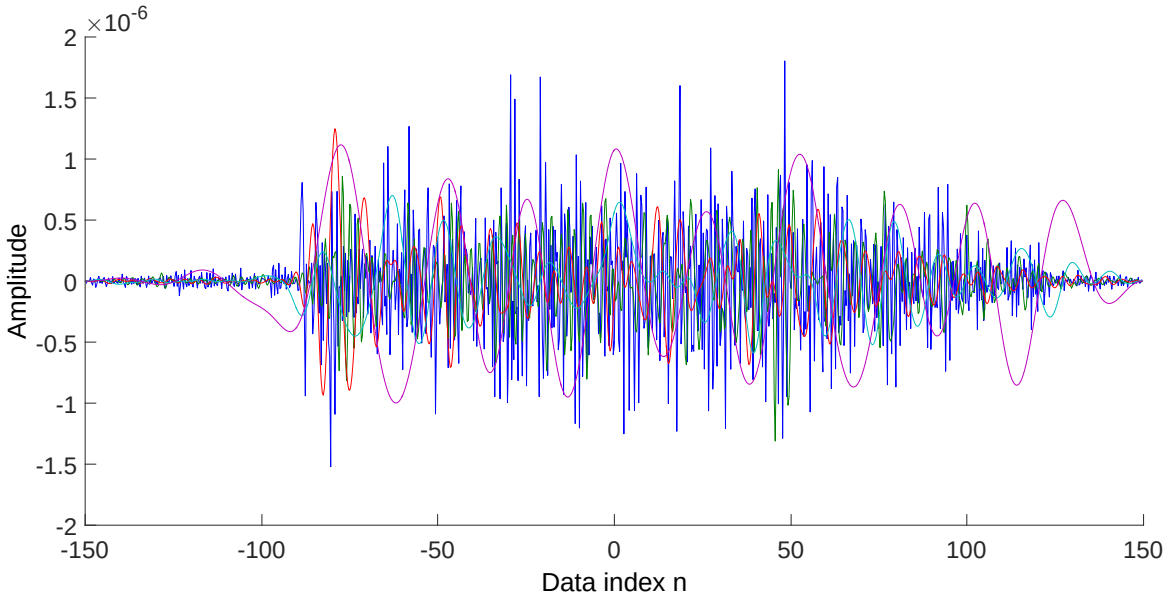


(a) Prediction EMD-CC State 1

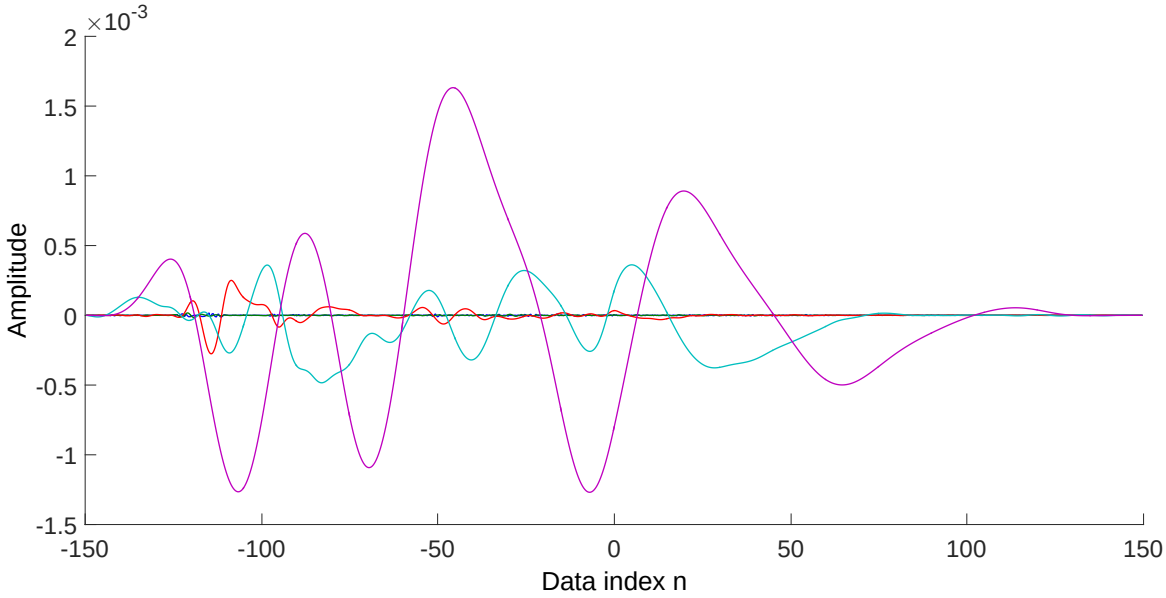


(b) Prediction EMD-CC State 2

Figure B.13: Graphical results of prediction EMD-CC



(a) Prediction EMD-CC State 3



(b) Prediction EMD-CC State 4

Figure B.14: Graphical results of prediction EMD-CC